



# **Previsão da procura de transportes públicos tendo em conta factores meteorológicos**

**Ricardo Fernandes Correia**

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Tânia Daniela Lopes da Rocha Fontes

Co-orientador: José Luís Cabral Moura Borges

28 de Fevereiro de 2020



# Abstract

Weather conditions have a major impact on citizens' daily mobility. Depending on weather conditions trips may be delayed, demand may be changed as well as the modal shift. These variations have a major impact on the use and operation of public transport, particularly in transport systems that operate close to capacity. However, the influence of weather conditions on transport demand is difficult to predict and quantify. For this purpose, an artificial neural network model – the Multilayer Perceptron – is used as a regression model to estimate the demand for urban public transport buses based on weather conditions. Hourly transit bus ridership and weather conditions were collected along a year from a medium-size european metropolitan area (Oporto, Portugal) and linked under the assumption that individuals choose the travel mode based on the weather conditions that are observed during the departure hour, the hour before and two hours before. The transit ridership data were also labelled according to the hour, day of the week, month, and whether there was a strike and/or holiday or not. The results demonstrate that it is possible to predict the demand of public transport buses using the weather conditions observed two hours before with low error for the entire network (7%) and relatively low error for each of the bus routes (37%). The use of weather conditions allows decreasing the error of the prediction by 22% for the entire network and by 15% for each route.





# Resumo

As condições meteorológicas têm um grande impacto na mobilidade diária dos cidadãos. Dependendo das condições meteorológicas, as viagens podem atrasar, a procura pode ser alterada, assim como o meio de transporte. Essas variações têm um grande impacto no uso e operação do transporte público, principalmente em sistemas de transporte que operam perto da capacidade máxima. No entanto, é difícil prever e quantificar a influência das condições meteorológicas na procura de transporte público urbano. Para esse fim, um modelo de rede neuronal artificial - o *Multilayer Perceptron* - é usado como modelo de regressão para estimar a procura nos autocarros públicos urbanos com base nas condições meteorológicas. Para isso, o número de passageiros de autocarros e as condições meteorológicas foram recolhidos ao longo de um ano para uma área metropolitana europeia de média dimensão (Porto, Portugal) e vinculados sob o pressuposto de que os utilizadores podem escolher efectuar uma viagem com base nas condições meteorológicas observadas durante a hora de partida, a hora anterior à partida e duas horas antes da partida. Os dados do número de passageiros também foram rotulados de acordo com a hora, dia da semana, mês e se houve uma greve e/ou feriado, ou não. Os resultados demonstram que é possível prever a procura de autocarros públicos urbanos usando as condições meteorológicas observadas duas horas antes da partida com baixo erro para toda a rede (7%) e erro relativamente baixo para cada uma das linhas de autocarros (37%). O uso de condições meteorológicas permite diminuir o erro de previsão em 22% para toda a rede e em 15% para cada uma das linhas.



# Agradecimentos

Gostaria de começar por agradecer à doutora Tânia Fontes e ao doutor José Luís Borges pelo apoio e orientação nesta dissertação. Uma palavra também de agradecimento ao INESC TEC e aos membros do CESE por me terem acolhido e ajudado sempre que necessário.

Aos meus pais por me guiarem e terem permitido, com o seu esforço e dedicação, que me tornasse uma pessoa melhor. A eles agradeço a amizade, o amor e tudo o que tenho de bom na vida.

Ao meu irmão pela amizade e por me mostrar que o empenho e a dedicação no que fazemos nos torna melhores e nos leva mais longe.

Por último gostaria de agradecer aos meus amigos e família que me acompanharam nesta jornada e que tenho a certeza que me acompanharão daqui para a frente.

Este trabalho foi financiado pelo FEDER (Fundo Europeu de Desenvolvimento Regional) através do Programa Operacional para a Competitividade e Internacionalização - Programa COMPETE 2020 e pelos Fundos Nacionais através da Fundação para a Ciência e a Tecnologia pelo projeto PTDC / ECI-TRA / 32053/2017 - POCI-01-0145-FEDER-032053.

Ricardo Fernandes Correia



*“Success is the ability to move from one failure to another without loss of enthusiasm”*

Winston Churchill



# Conteúdo

<b>Abstract</b>	<b>i</b>
<b>Resumo</b>	<b>iii</b>
<b>Agradecimentos</b>	<b>v</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Objectivos . . . . .	2
1.3 Estrutura da dissertação . . . . .	2
<b>2 Revisão Bibliográfica</b>	<b>5</b>
2.1 Influência das condições meteorológicas na mobilidade . . . . .	6
2.2 Influência das condições meteorológicas nos transportes públicos . . . . .	6
2.3 Dados e métodos de análise . . . . .	8
2.4 Modelos de previsão . . . . .	10
<b>3 Redes Neurais Artificiais</b>	<b>13</b>
3.1 Neurónio artificial . . . . .	14
3.2 Algoritmos de optimização . . . . .	15
3.3 Definição de hiperparâmetros . . . . .	16
3.4 Medidas de <i>performance</i> . . . . .	17
<b>4 Mobilidade Urbana na Área Metropolitana do Porto</b>	<b>19</b>
4.1 Área metropolitana do Porto . . . . .	19
4.2 Sistema Andante . . . . .	20
4.2.1 Descrição do serviço de transportes na área metropolitana do Porto . . . . .	20
4.3 Características da mobilidade na área metropolitana do Porto . . . . .	23
<b>5 Metodologia</b>	<b>25</b>
5.1 Recolha de dados . . . . .	27
5.2 Definição de grupos . . . . .	28
5.3 Definição de cenários para escolha das variáveis de entrada no modelo . . . . .	28
5.4 Modelação . . . . .	32
5.4.1 O modelo . . . . .	32
5.4.2 Aprendizagem . . . . .	33
5.4.3 Optimização . . . . .	34
5.4.4 Indicadores de <i>performance</i> . . . . .	35

<b>6</b>	<b>Resultados</b>	<b>37</b>
6.1	Avaliação da <i>performance</i> do modelo . . . . .	37
6.1.1	Avaliação dos modelos de previsão da procura em toda a rede . . . . .	38
6.1.2	Avaliação dos modelos de previsão da procura por linha . . . . .	39
6.2	Resultados dos testes para as melhores redes neuronais . . . . .	39
6.2.1	Avaliação da <i>performance</i> por estação do ano . . . . .	40
6.2.2	Avaliação da <i>performance</i> por dia da semana . . . . .	40
6.2.3	Avaliação da procura para diferentes tipos de dias . . . . .	41
<b>7</b>	<b>Ferramenta de previsão</b>	<b>43</b>
<b>8</b>	<b>Conclusões</b>	<b>45</b>
8.1	Trabalho futuro . . . . .	47
<b>A</b>	<b>Análise dos dados recolhidos</b>	<b>49</b>
<b>B</b>	<b>Resultados do teste Kolmogorov-Smirnov</b>	<b>55</b>
<b>C</b>	<b>Resultados gráficos da avaliação da procura para diferentes tipos de dias</b>	<b>57</b>
<b>D</b>	<b>Código da ferramenta</b>	<b>65</b>



# Lista de Figuras

3.1	Rede com uma camada escondida (a) e múltiplas camadas (b) (Fonte: Nielsen and A. (2019)) . . . . .	14
3.2	Exemplo de um neurónio artificial (Fonte: Haykin et al. (2009)) . . . . .	15
4.1	Exemplo de uma estação (a) e paragem (b) de metro da AMP (Flickr (2019)). . .	21
4.2	Exemplo de estação de autocarro (a) e paragens (b e c) de autocarro da AMP (Porto. (2018); JPN (2019)) . . . . .	21
4.3	Mapa do zonamento Andante (AMP (2019)) . . . . .	22
4.4	Mapa de zonas de utilização do sistema Andante na Área Metropolitana do Porto (TIP - Transportes Intermodais do Porto (2013)) . . . . .	24
5.1	Diagrama da metodologia utilizada para definição do modelo da ANN. . . . .	26
5.2	Gráfico (a) e função (b) da função de ativação Leaky ReLU. . . . .	33
5.3	Momento de primeiro e segundo grau . . . . .	34
5.4	Actualização de parâmetros . . . . .	35
7.1	Interface da ferramenta . . . . .	44
7.2	Diagrama do funcionamento da ferramenta. . . . .	44
A.1	Número de validações mensal no sistema de transportes públicos andante (STCP) em 2013. . . . .	49
A.2	Valores de temperatura (°C) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990 . . . . .	50
A.3	Valores de velocidade do vento (Km/h) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990 . . . . .	51
A.4	Percentagem de humidade relativa (%) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990 . . . . .	51
A.5	Valores de pressão atmosférica (mb) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990 . . . . .	52
A.6	Procura média por grupo de utilizador e hora durante o ano de 2013 . . . . .	52
C.1	Procura total para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia da semana: Terça-feira, Nebulosidade: céu limpo . . . . .	57
C.2	Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia da semana: Terça-feira, Nebulosidade: céu limpo	58
C.3	Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Fim de semana: Sábado, Nebulosidade: chuva . . . . .	59
C.4	Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Fim de semana: Sábado, Nebulosidade: chuva . .	60

C.5	Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Feriado: Quarta-feira, Nebulosidade: algumas nuvens	61
C.6	Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Feriado: Quarta-feira, Nebulosidade: algumas nuvens . . . . .	62
C.7	Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Greve: Terça-feira, Nebulosidade: nublado . . . . .	63
C.8	Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Greve: Terça-feira, Nebulosidade: nublado	64

# Lista de Tabelas

2.1	Métodos e variáveis usadas em estudos anteriores . . . . .	9
5.1	Exemplo da estrutura de dados de validações do sistema Andante . . . . .	27
5.2	Descrição das variáveis usadas nas definições dos cenários . . . . .	28
5.3	Cenários aplicados para a procura total e para a procura por linha de operação . .	30
5.4	Correlação de Spearman aplicado à procura da rede de transportes públicos (auto-carros) . . . . .	31
6.1	Resultados da pesquisa aleatória para previsão da procura total. . . . .	37
6.2	Resultados da pesquisa aleatória para previsão da procura por linha. . . . .	37
6.3	Resultados da validação cruzada para a procura total da rede de autocarros. . . .	38
6.4	Resultados da validação cruzada para a procura por linha de operação de autocarros	39
6.5	Resultados da previsão da procura em diferentes estações do ano. . . . .	40
6.6	Resultados da previsão da procura de autocarros em diferentes dias da semana. .	41
6.7	Resultados da previsão para diferentes tipos de dias. . . . .	42
A.1	Valores estatísticos das variáveis meteorológicas . . . . .	53
B.1	Teste de Kolmogorov-Smirnov . . . . .	56



# Abreviaturas e Símbolos

ACE	Agrupamento Complementar de Empresas
ADAM	Estimação Adaptativa de Momento
AML	Área Metropolitana de Lisboa
AMP	Área Metropolitana do Porto
ANN	Artificial Neural Network
ANOVA	Análise de Variância
API	Interface de Programação de Aplicação
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CP	Comboios de Portugal
CPU	Unidade Central de Processamento
DNN	Deep Neural Network
EU	União Europeia
EU-28	Vinte e Oito Estados Membros da União Europeia
GPU	Unidade de Processamento Gráfico
ID	Identificador Único
IPMA	Instituto Português do Mar e da Atmosfera
JSON	JavaScript Object Notation
MAE	Erro Médio Absoluto
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Erro Médio Quadrático
NUT	Nomenclatura das Unidades Territoriais para Fins Estatísticos
OLS	Ordinary Least Squares
OMM	Organização Meteorológica Mundial
RMSE	Raiz do Erro Médio Quadrático
SARIMA	Seasonal Autoregressive Integrated Moving Average
SGD	Gradiente Descendente Estocástico
SIA	Sistema Intermodal Andante
TIP	Transportes Intermodais do Porto



# Capítulo 1

## Introdução

As condições meteorológicas têm um grande impacto no dia a dia dos cidadãos, especialmente quando se trata de dias em que as condições meteorológicas não são favoráveis a viagens, como por exemplo a ocorrência de chuva, vagas de frio ou ondas de calor (Shaaban and Muley (2016); Hjorthol (2013)). Embora estas condições meteorológicas menos favoráveis possam ocorrer durante todo o ano, são mais notórias para a população em determinados períodos do ano, tal como no inverno e no verão. Assim, a influência na tomada de decisão num contexto de mobilidade, em particular de um utilizador de transportes públicos, é por essa razão maior nesses períodos. Dependendo do perfil de utilizador (e.x. idoso ou doente), a ocorrência de condições meteorológicas severas pode levar um utilizador a alterar o meio de transporte, ou em situações mais extremas, a adiar ou mesmo a cancelar uma viagem. Porém, embora a avaliação destes impactos seja algo intuitivo, a sua previsão e quantificação são mais complexas.

A análise de estudos realizados demonstra que o principal foco, e consequentemente os modelos de previsão criados, tem sido na previsão da chegada de autocarros às respectivas paragens. Ao mesmo tempo os modelos desenvolvidos para a previsão da procura têm-se focado essencialmente na previsão da procura a curto prazo (Xue et al. (2015) e Ma et al. (2014)), recorrendo a sistemas automáticos de validações que fornecem a procura em tempo real. Existe assim uma falha na literatura no que toca à previsão da procura nos transportes públicos tendo em contas as condições meteorológicas.

De modo a colmatar a falta de modelos de previsão nos autocarros e ao mesmo tempo propor uma nova solução, foi desenvolvida uma ferramenta de previsão recorrendo a redes neuronais, algo que até ao momento não foi usado, de modo a prever a procura total e por linha da rede de autocarros tendo em conta as condições meteorológicas.

### 1.1 Motivação

No mundo actual um dos grandes objectivos da gestão de qualquer empresa é tentar oferecer a melhor qualidade de serviço possível com um custo menor. Nas empresas de transporte público isso não é diferente.

A qualidade de serviço pode ser analisada segundo dois pontos de vista, o primeiro do ponto de vista do utilizador e o segundo do ponto de vista da organização. Estas duas visões são bastante distintas. Enquanto o primeiro responde a perguntas como: "o meio de transporte é confortável e seguro?", o segundo responde a perguntas como: "o meio de transporte é eficiente ou financeiramente sustentável?".

A alteração da procura nos transportes públicos pode criar por sua vez problemas na gestão das operações das empresas prestadoras de serviços de transporte, bem como constrangimentos nos serviços prestados, ou seja diminuição na qualidade de serviço, podendo os recursos disponíveis (e.x. veículos e motoristas) não serem adequados às necessidades observadas. Assim, diversas condições externas podem gerar alterações na qualidade de serviço e na gestão dos recursos disponíveis.

## 1.2 Objectivos

O objectivo deste trabalho consiste no desenvolvimento de uma ferramenta de previsão da procura nos transportes públicos, mais especificamente nos autocarros, tendo em conta as condições meteorológicas.

Para isso foram definidas as seguintes questões:

- Quais são as variáveis meteorológicas que mais influenciam a procura de autocarros públicos urbanos?
- O uso de variáveis meteorológicas ajuda a melhorar a exactidão de um modelo de previsão da procura de autocarros públicos urbanos?
- A previsão das viagens em autocarros públicos urbanos aumenta com o uso das condições meteorológicas observadas duas horas antes das viagens?

De modo a atingir os objectivos delineados, será utilizado um modelo para estimar a procura nos autocarros públicos urbanos. A performance do modelo será avaliada usando diferentes conjuntos de dados recolhidos ao longo de um ano para um caso de estudo real, a Área Metropolitana do Porto. A robustez do modelo será analisada para diferentes períodos do dia, estações do ano, grupos de utilizadores e condições meteorológicas. Os resultados deste modelo serão integrados, juntamente com dados recolhidos através de API's, numa ferramenta que permitirá prever a procura de autocarros públicos urbanos.

## 1.3 Estrutura da dissertação

A presente dissertação encontra-se dividida em 8 capítulos.

O Capítulo 1 aborda o enquadramento do trabalho, as suas principais motivações e objectivos e descreve a estrutura do trabalho.



No Capítulo 2 é apresentado o estado da arte e revisão bibliográfica, onde são apresentadas as influências das condições meteorológicas na mobilidade e nos transportes públicos. É ao mesmo tempo efectuada uma análise dos estudos realizados, até ao momento, e apresentados modelos de previsão da procura nos transportes públicos.

No Capítulo 3 é apresentada uma revisão bibliográfica sobre redes neuronais.

No Capítulo 4 é elaborada uma análise sobre a mobilidade urbana na área metropolitana do Porto. Neste capítulo é caracterizado o sistema de validação dos transportes públicos e os serviços oferecidos pelos operadores de transporte publico desta área metropolitana.

No Capítulo 5 é apresentada a metodologia usada para a resolução do problema, designadamente os dados, os cenários e o modelo usado, e no Capítulo 6 apresentam-se e analisam-se os resultados obtidos com o desenvolvimento desta metodologia para um conjunto de testes.

Por fim, o Capítulo 7 descreve o funcionamento da ferramenta final desenvolvida e o Capítulo 8 as conclusões finais do trabalho, bem como alguns pontos para possível trabalho futuro.



## Capítulo 2

# Revisão Bibliográfica

A mobilidade é caracterizada pela facilidade de deslocamento de bens e pessoas podendo ser realizada usando diversos meios de transporte: motorizados e não motorizados. Assim sendo, a mobilidade é essencial para o bom funcionamento e desenvolvimento de uma sociedade, permitindo que se realizem mais facilmente actividades económicas e sociais (Rodrigue et al. (2017)). Este deslocamento é essencialmente caracterizado em dois tipos: transporte de mercadorias e transporte de pessoas.

No que respeita à mobilidade de pessoas nos países pertencentes à União Europeia (EU), denota-se uma clara preferência no uso de veículo particular como meio de transporte. Este meio de transporte é usado diariamente por 83,1% da população. Em contrapartida, os meios de transporte metro e autocarro representam 9,2% sendo seguidos por comboio com 7,7% (Eurostat).

Dentro dos 28 estados membros da União Europeia existem algumas assimetrias no que diz respeito às preferências dos cidadãos. A Hungria e a República Checa foram os únicos Estados-Membros em que as percentagens no uso de veículos particulares se situam abaixo de três quartos. Nestes países o uso de autocarro e metro tem uma maior representação, com 22,3% e 17,3% respectivamente, em relação aos restantes estados membros (Eurostat).

É amplamente reconhecido que a procura nos transportes, sejam estes privados ou públicos, é o produto das preferências individuais de cada cidadão. Diversos factores influenciam as escolhas de cada utilizador de transportes, desde: (i) as características pessoais do utilizador (ex: género, idade, ocupação, rendimento, e ter ou não veículo), (ii) a disponibilidade de transportes públicos (ex: meio de transporte, tempo de viagem, custo e segurança), (iii) o objectivo da viagem (ex: lazer ou laboral), (iv) o dia e o período do dia (ex: fim de semana vs. semana, hora de ponta vs. hora de vazio) e (v) as condições meteorológicas (ex: precipitação, altas ou baixas temperaturas e ventos fortes).

Este capítulo apresenta uma análise dos estudos realizados, sendo analisadas a influência das condições meteorológicas na mobilidade (secção 2.1), a influência das condições meteorológicas nos transportes públicos (secção 2.2), os métodos utilizados nos estudos realizados (secção 2.3) e os modelos de previsão até agora criados (secção 2.4).

## 2.1 Influência das condições meteorológicas na mobilidade

Diversos autores estudaram a influência das condições meteorológicas na mobilidade dos cidadãos (Sabir et al. (2010); Shaaban and Muley (2016)). Estas condições afectam de forma diferente os diversos meios de transporte (Cools et al. (2010); Singhal et al. (2014)) e tipos de utilizadores (Mahrsi (2014); Hjorthol (2013)).

No que respeita ao fluxo pedestre, num estudo realizado por de Montigny et al. (2011) em nove cidades europeias concluiu-se que o nível de precipitação, a temperatura e a luz solar têm uma influência significativa em flutuações no fluxo de pedestres. Segundo este estudo, uma variação positiva de 5°C em relação à temperatura média do local provocou um aumento de 14% na média dos cidadãos que se deslocam a pé. Mais ainda, uma mudança de um dia com neve para um dia sem precipitação gerou um aumento de 23% de pedestres. Shaaban and Muley (2016), num estudo realizado no Catar, constatou que o fluxo de pedestres tem uma relação logarítmica-linear com as condições meteorológicas e que a temperatura é o parâmetro que mais influencia o fluxo de pedestres.

Em relação ao uso de bicicleta, em condições de tempo chuvoso ou na presença de chuva durante as três horas anteriores, ciclistas tendem a adiar ou mesmo cancelar as suas viagens. Estudos realizados por Fields et al. (2012) e Parkin et al. (2008) revelam que o volume de ciclistas tende a aumentar com o aumento de temperatura e a diminuição da precipitação. Ao mesmo tempo, Lewin (2011) verifica que os efeitos da temperatura e da precipitação no fluxo de ciclistas não é explicado por uma regressão linear, podendo existir uma diminuição quer em condições com altas temperaturas quer em condições com baixas temperaturas. Nos casos de ocorrência de precipitação a diminuição de fluxo de ciclistas na primavera é de 8% e no verão de 13%.

No uso de veículo privado tal influência não é tão denotada devido às suas características, tais como o isolamento das condições meteorológicas. Os passageiros não são directamente influenciados pelas condições exteriores, sendo que as condições meteorológicas apenas influenciam a rapidez da viagem realizada. Segundo Hooper et al. (2014), no uso de carro, a precipitação influencia a velocidade, o fluxo do trânsito e consequentemente o tempo de viagem.

O uso dos transportes públicos é dos meios de transporte mais influenciados pelas condições meteorológicas, visto que os seus utilizadores são directamente afectados quando se dirigem ou aguardam numa paragem ou estação por um transporte público. Estas influências são analisadas em detalhe na secção seguinte (2.2).

## 2.2 Influência das condições meteorológicas nos transportes públicos

A influência da meteorologia nos transportes públicos não é um tema recente e tem gerado um crescente interesse no seu estudo e análise. O estudo deste tema tem-se focado essencialmente nos autocarros (Tao et al. (2016), Stover and McCormack (2015)) e no metro (Li et al. (2018); Zhou et al. (2017)) como meio de transporte, utilizando diferentes métodos de análise em diferentes localizações e sociedades.

Diversos estudos demonstram a existência de uma correlação directa entre a variação das condições meteorológicas e a procura gerada nos serviços de transporte (Singhal et al. (2014), Guo et al. (2008)). Outros autores (Guo et al. (2007); Kalkstein et al. (2009)) demonstram com os seus resultados que a utilização de autocarros e metro por parte dos utentes é negativamente afectada com elevada precipitação.

A meteorologia, como um dos principais factores influenciadores da mobilidade, intervém nas decisões dos utilizadores e nos seus hábitos. Guo et al. (2008) aferiu que a utilização dos transportes públicos é directamente influenciada pela meteorologia, sendo esta influência negativa quando as condições não são favoráveis e vice-versa. A temperatura, a humidade relativa, a precipitação e a velocidade do vento são as principais variáveis que mais contribuem para estas variações (Zhou et al. (2017); Guo et al. (2008)), sendo que a severidade do seu impacto depende de variável para variável, do meio de transporte e do dia da semana. Segundo Cools et al. (2010) as condições meteorológicas tendem a determinar o comportamento dos utilizadores na escolha de viajar ou não, bem como na escolha de destino, meio de transporte e rota a seguir. Este concluiu igualmente, que condições adversas têm um maior impacto nas viagens ocasionais.

Sabir et al. (2010) afere que em geral a meteorologia tem influência na procura gerada nos transportes públicos e, mais importante ainda, que as condições meteorológicas afectam de forma diferente cada um dos meios de transporte público.

Singhal et al. (2014) averiguou a correlação, com uma periodicidade horária, entre as condições meteorológicas e as viagens realizadas no metro de Nova Iorque (EUA) à semana e ao fim de semana. O seu estudo revelou que diversas variáveis, tais como a precipitação e neve, elevada velocidade do vento, contribuíam negativamente na procura deste meio de transporte. Do mesmo modo Costa et al. (2017) concluíram que a temperatura é uma das variáveis que mais afecta os utentes, ao contrário da pressão atmosférica que aparenta não causar nenhum efeito negativo.

As condições meteorológicas aparentam afectar de forma diferente o uso dos transportes públicos, dependendo do dia da semana. Verifica-se que condições adversas induzem uma maior variação na procura de transportes públicos aos fins-de-semana, onde a precipitação, a elevada velocidade do vento e baixas temperaturas afectam severamente a sua utilização. Segundo Li et al. (2018) esta diferença na procura deve-se ao propósito das viagens realizadas nesses espaços de tempo, sendo que durante a semana a maioria é realizada com o objectivo de ir para o local de trabalho, enquanto que ao fim de semana é mais em torno do lazer. Em Guipúscoa (Espanha) num estudo, realizado por Arana et al. (2014), foram usados dados diários das variáveis meteorológicas e de viagens realizadas à semana e ao fim-de-semana sendo concluído que a elevada velocidade do vento e precipitação causaram uma diminuição acentuada das viagens realizadas, em sentido oposto um aumento de temperatura tende a aumentar essa procura.

Por outro lado, os efeitos das condições meteorológicas tendem a variar caso se verifiquem grandes eventos tais como feriados nacionais e datas comemorativas. De notar que, como mencionado por Kalkstein et al. (2009), a estação do ano apenas desempenha um pequeno papel no impacto das condições meteorológicas na procura dos transportes públicos, sendo que um determinado evento meteorológico provoca respostas semelhantes em diferentes estações do ano.

Cools et al. (2010) num estudo realizado em Uccle (Bélgica), concluiu que os movimentos pendulares são os menos afectados por estas mudanças meteorológicas, principalmente no que toca a cancelar a viagem.

Numa análise mais aprofundada Yagi and Mohammadian (2008) realçou a importância da distância da viagem, género, idade e possuir ou não carta de condução como factores que contribuem na escolha de tipo de transporte a utilizar, seja este privado ou público. Este pensamento mais voltado para as variáveis sócio-económicas despoletou em alguns autores a necessidade de estudar as influências das condições meteorológicas em diferentes estratos sociais e idades (Mahrsi (2014)). Hjorthol (2013) num estudo realizado na Noruega, concluiu que condições adversas têm maior impacto nos passageiros seniores, com idade superior a 75 anos, no inverno comparativamente ao verão, este facto segundo o autor deve-se à sua diminuição da mobilidade e à sua maior fragilidade às condições do meio ambiente.

### 2.3 Dados e métodos de análise

A recolha de dados com o objectivo de encontrar relações entre as condições meteorológicas e a sua influência nos transportes públicos tem sido essencialmente realizada de duas formas distintas: (i) através de questionários a utilizadores, ou (ii) através de recolha passiva de informação recorrendo a sistemas de validação automáticos. Os primeiros tendem a ser pouco exactos e dispendiosos, quer em tempo quer em recursos, tendo como principal vantagem a recolha de informação mais pessoal do utilizador, tal como a idade e ocupação. Os segundos, por sua vez são mais vantajosos pois permitem a obtenção de grandes quantidades de dados com poucos erros e em tempo diminuto. Tendo em conta as vantagens e desvantagens de cada um destes métodos, os autores tendem a preferir o segundo método de recolha de dados para realizar os seus estudos.

Na Tabela 2.1 estão representados estudos que analisaram a influência das condições meteorológicas na mobilidade dos cidadãos, quer em meios de transporte público quer em meios de transporte privado. Estes estudos foram realizados em diferentes cidades, de diferentes dimensões e características sócio-económicas e culturais distintas.

No que diz respeito às variáveis utilizadas nos estudos, todos estes artigos analisaram a influência da temperatura e apenas um não analisou a influência da precipitação. A velocidade do vento aparenta ser ao mesmo tempo uma variável de grande interesse estando presente em cerca de 70% dos estudos. A variável menos utilizada foi a pressão atmosférica, sendo apenas analisada num estudo. Acerca da periodicidade das variáveis utilizadas, os autores optam por periodicidades diárias e horárias, havendo uma ligeira preferência pela periodicidade diária, cerca de 55%.

Como pode ser observado na Tabela 2.1, os estudos são predominantemente realizados em transportes públicos, nomeadamente no Metro e autocarros.

Tabela 2.1: Métodos e variáveis usadas em estudos anteriores

Referência	Cidade	Método	Periodicidade	Tipo de Transporte			Variáveis Meteorológicas					Principais Conclusões	
				B	M	Bic	F	T (°C)	R (mm)	RH (%)	WS (km/h)		SNW (mm)
Tao et al. (2018)	Brisbane, Austrália	Arimax e Sarimax	1 H	x				x	x	x			Alterações das condições meteorológicas podem gerar flutuações na procura de autocarros.
Li et al. (2018)	Nanjing, China	Média Móvel de 13 termos	Diária		x			x	x		x		A precipitação aparenta ter mais influência na procura do que a temperatura.
Tao et al. (2016)	Brisbane, Austrália	ANOVA	Diária	x				x	x	x			O metro é menos influenciado do que os autocarros.
Zhou et al. (2017)	Shenzhen, China	Regressão Multivariate	1 H	x	x			x	x	x			O aumento da humidade, vento e chuva gera uma diminuição na procura nos transportes públicos.
Costa et al. (2017)	Porto, Portugal	Coefficiente de correlação de Spearman	1 H	x	x			x		x	x	x	A pressão atmosférica não aparenta influenciar a procura nos transportes públicos ao contrário da temperatura.
Creemers et al. (2015)	Holanda	Regressão Linear	1 H					x	x				A temperatura, precipitação e nevoeiro aparentam influenciar a tomada de decisão dos utilizadores.
Singhal et al. (2014)	New York, EUA	Método Mínimos Quadrados	1 H	x	x			x	x		x		A precipitação e a neve influenciam a procura nos transportes públicos, principalmente aos fins de semana.
Stover and McCormack (2015)	Pierce County, EUA	Método Mínimos Quadrados	Diária	x				x	x		x		As condições meteorológicas têm maior influência no inverno do que nas outras estações do ano.
de Montigny et al. (2011)	Diversas Cidades	Regressão Quasi-Poisson	Diária				x	x		x			A temperatura é a variável que mais influencia o fluxo de pedestres.
Arana et al. (2014)	Guipúscoa, Espanha	Regressão Multivariável	Diária	x				x	x	x			A precipitação e vento têm maior influência ao fim de semana, na procura de transporte público, tendo a temperatura amena uma influência contrária.
Schmiedeskamp and Zhao (2016)	Seattle, EUA	Distribuição binomial negativa	Diária			x		x	x				A temperatura mínima reportada pelos serviços de meteorologia aparentam ter maior influência na tomada de decisão do que a temperatura máxima reportada.

Nota:  
Tipo de transporte: B - Autocarro; M - Metro; Bic - Bicicleta; F - A Pé;  
Variáveis Meteorológicas: T - Temperatura; R - Precipitação; RH - Humidade Relativa; WS - Velocidade do Vento; SNW - Neve; P - Pressão Atmosférica;

Os métodos implementados, para encontrar relações entre as variáveis meteorológicas e a procura nos transportes públicos, são diversificadas e variam de autor para autor. Singhal et al. (2014), num estudo realizado em Nova Iorque (EUA) com dados de viagens de autocarro e metro, com uma periodicidade horária, referentes aos anos 2010 e 2011 e dados meteorológicos com a mesma periodicidade, aplicou modelos de mínimos quadrados (OLS) utilizando valores diários, horários. Foram, do mesmo modo, aplicados modelos para dias da semana e fins de semana.

A utilização do método de Mínimos Quadrados permite otimizar os parâmetros de um modelo de modo a minimizar a soma dos erros ao quadrado. Este método, encontra mínimos globais se o modelo utilizado for linear. Sendo vantajoso quando se conhece ou se tem alguma intuição sobre a forma analítica do modelo original dos dados. A principal vantagem desta técnica é que permite estimar o valor esperado de uma dada variável tendo como entrada uma ou mais variáveis. A variável dependente deve ser contínua, ou próxima de contínua, caso não o seja este método não pode ser implementado. Nos casos apresentados anteriormente a variável dependente é muito próxima de contínua o que permite a sua implementação. É de realçar que embora as variáveis independentes possam não ser contínuas, este modelo beneficia caso estas o sejam, pois, a variável dependente é directamente afectada por estas, obtendo-se assim um modelo mais aproximado da realidade (Rao (2008)).

Li et al. (2018) recorreu a uma média móvel de treze termos conjuntamente com uma análise de variância, alargando o conceito inicialmente introduzido por Kalkstein et al. (2009). Esta conjugação de técnicas torna-se especialmente vantajosa na análise da correlação entre as condições meteorológicas e procura nos transportes públicos num determinado dia da semana pois permite aferir se a flutuação na procura se deve realmente às condições desse dia ou se houve outra variável externa a induzir esta flutuação.

Tao et al. (2016) realizou o seu estudo recorrendo a uma Análise de Variância (ANOVA) e testes  $t$  para analisar a influência das condições meteorológicas nos padrões geográficos de uso nos autocarros. O método ANOVA torna-se especialmente vantajoso quando o número de populações é superior a três permitindo a análise de vários factores ao mesmo tempo.

## 2.4 Modelos de previsão

Embora a previsão da procura em transportes públicos, nomeadamente no autocarros, tenha atraído maior atenção nos últimos anos, poucos estudos e modelos de previsão foram desenvolvidos. Uma maior quantidade de modelos têm-se focado essencialmente na previsão da chegada de autocarros às suas paragens em função das condições meteorológicas e das condições de trânsito.

Cyril et al. (2018), com dados anuais de viagens realizadas durante os anos de 2010, 2011 e 2012 em Kerala (Índia), criou um modelo de previsão da procura anual utilizando como base um modelo ARIMA. Este modelo realizou a previsão da procura nos autocarros em Kerala para o ano de 2013 tendo obtido bons resultados em comparação com a procura real registada. O modelo criado por Cyril et al. (2018) provou ser preciso na previsão da procura de curto prazo entre distritos de Kerala altamente dependentes uns dos outros.



Xue et al. (2015) e Ma et al. (2014) criaram modelos de previsão a curto prazo (1h) recorrendo a modelos ARMA, SARIMA, and ARIMA que posteriormente, devido às características de cada um destes três modelos, os resultados são processados por algoritmos que têm como objectivo maximizar as vantagens de cada um destes modelos e otimizar a interacção entre estes. A ideia central destes algoritmos é estimar dinamicamente as combinações optimizadas dos modelos de previsão, usando dados em tempo real.

O uso de redes neuronais, como modelo de previsão, não aparenta ter despertado interesse nos estudiosos desta área específica. Mukai and Yoden (2012) foram uns dos poucos a usar redes neuronais como modelo de previsão da procura de transporte, mais especificamente nos táxis. Embora estes modelos de previsão não sejam muito usados, as redes neuronais artificiais (ANN) têm vantagem sobre os métodos convencionais de previsão, como a regressão, para simular sistemas complexos. Ao contrário da regressão, que precisa de uma função explicitamente definida para relacionar a entrada e a saída, a ANN pode aproximar uma função e associar a entrada à saída específica após ser adequadamente treinada pela entrada e pelos dados de saída correspondentes. Essa capacidade de associação é extremamente útil quando o sistema é bastante complexo (como o sistema de transportes públicos) e/ou é impossível especificar uma função para representar com precisão as relações entre a entrada e a saída. Além disso, a ANN não exige que as variáveis de entrada sejam independentes uma da outra, o que poderia economizar um esforço substancial na análise de correlação entre os dados.



## Capítulo 3

# Redes Neurais Artificiais

As redes neuronais artificiais (ANN's) têm a sua origem no estudos sobre a estrutura do cérebro humano, na tentativa de reproduzir a sua forma inteligente de processar informação (Baughman and Liu (2014)). Uma ANN é por definição uma rede massivamente paralela de elementos interconectados e as suas organizações hierárquicas estão preparadas para conjugar objectos do mundo real, da mesma forma que um sistema nervoso biológico o faz. Por outras palavras, as ANN's são sistemas adaptativos desenhados com o intuito de simular a forma como os neurónios humanos estão ligados.

Uma ANN é composta por várias unidades de processamento: os neurónios e as conexões que os ligam. O comportamento “inteligente” de uma ANN tem origem nas interações entre as unidades de processamento da rede, que detêm um processo de aprendizagem realizado através de um algoritmo. Assim, à semelhança do cérebro humano, as ANN detêm a capacidade de aprender e agir perante as mais diversas situações, bem como adquirir informação através da experiência e da observação. As ANN podem ainda modificar a sua própria tipologia tal qual como os neurónios no cérebro (Haykin et al. (2009)).

Numa ANN os neurónios estão organizados em camadas, existindo pelo menos duas camadas, entrada e saída. As camadas são classificadas em um de três tipos: camada de entrada; camada(s) escondida(s); camada de saída. Uma ANN pode ter uma ou mais camadas escondidas, sendo o segundo designado *Deep Neural Network* (DNN). Na Figura 3.1 pode ser observado cada um destes dois tipos de redes.

Uma ANN pode ser classificada em um de dois tipos de rede dependendo de como os neurónios se encontram ligados entre as camadas:

- Rede feedforward: neste tipo de rede os dados circulam numa única direcção, iniciando o seu trajecto na camada de entrada e terminando na camada de saída. Este tipo de rede é um mapeamento estático entre as camadas de entrada e saída, que permite reflectir uma relação linear simples ou uma relação não linear;
- Rede feedback: neste tipo de rede as saídas são utilizadas também para alimentar as próprias entradas. Devido à sua configuração pode adquirir memória a curto prazo.

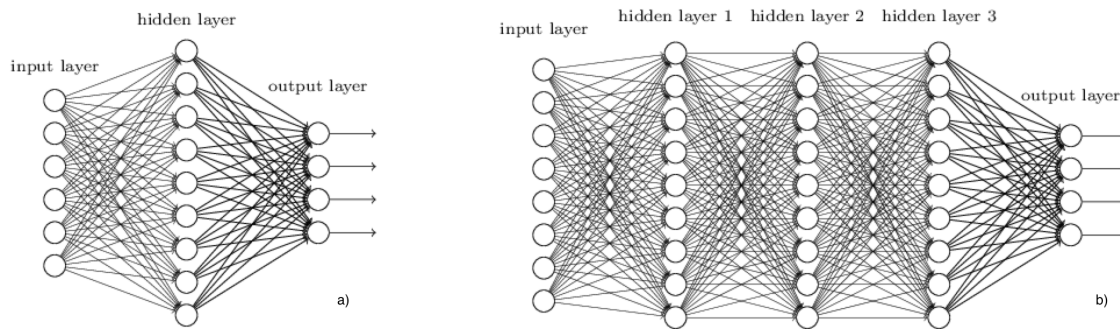


Figura 3.1: Rede com uma camada escondida (a) e múltiplas camadas (b) (Fonte: Nielsen and A. (2019))

O poder de abstracção das ANN deve-se à sua estrutura paralela e à capacidade de aprendizagem das mesmas. A estrutura paralela advém da existência de muitos neurónios ligados numa estrutura de pesos de conexão e com fácil adaptação. A estrutura paralela é desejável uma vez que permite tolerância à falha, pois se algum neurónio, ou neurónios, falhar, os efeitos na rede não serão significativos para o desempenho desta.

Segundo Haykin et al. (2009) as ANN potenciam certas propriedades tais como:

- Não linearidade: um neurónio artificial pode ser linear ou não linear. Se existir pelo menos um neurónio não linear, a ANN poderá ter a capacidade de mapear uma função não-linear, ou seja, uma ANN composta por interconexões de neurónios não-lineares, é em si não-linear;
- Mapeamento de entradas-saídas: ilustra o que acontece no método de aprendizagem supervisionada, em que a rede é treinada repetidamente;
- Adaptabilidade: as redes neuronais têm a capacidade interna de adaptarem os seus pesos de ligação às mudanças ao seu redor;
- Tolerância a falhas: as redes neuronais têm a capacidade de continuar a executar a sua função mesmo quando na fase de treino, um ou mais neurónios, deixam de poder actualizar os seus valores de pesos das conexões a outros neurónios.

### 3.1 Neurónio artificial

A unidade base de processamento de uma ANN é o neurónio artificial. O neurónio artificial, como apresentado no início do capítulo 3, é inspirado no neurónio do cérebro humano e como tal pode ter uma ou mais entradas e uma única saída.

Uma das propriedades mais importantes de um neurónio artificial é o vector com os pesos das entradas. Os pesos deste vector são representados pela letra  $W$  e representam a importância que uma determinada ligação de entrada possui em relação a um determinado neurónio. O valor de cada um destes pesos é actualizado a cada ciclo de aprendizagem da rede.

Cada neurónio tem a si associada uma função de activação. De um modo geral cada camada da ANN tem apenas um tipo de função de activação e pode ser um de três tipos: linear, limiar e não-linear.

- Limiar: este tipo de função de activação tem apenas dois valores possíveis, 0 ou 1. A função limiar é muito usada na camada de saída, de redes neuronais, em problemas de classificação;
- Linear: este tipo de função de activação usa os valores de entrada multiplicados pelo vector de pesos do neurónio, obtendo assim uma saída proporcional à entrada. A função linear é muito usada na camada de saída, de redes neuronais, em problemas de regressão;
- Não-linear: este tipo de função de activação permite que o modelo crie mapeamentos complexos entre as entradas e saídas da ANN, essenciais para aprender e modelar dados complexos. Actualmente o uso deste tipo de função de activação é muito usado em todo o tipo de redes neuronais, principalmente nas camadas escondidas, pois em termos teóricos quase qualquer processo pode ser representado por uma função não-linear.

Na Figura 3.2 é apresentada a estrutura base de um neurónio. Nesta podemos notar os diversos componentes do neurónio, tal como as entradas, a saída, os pesos das ligações e a função de activação.

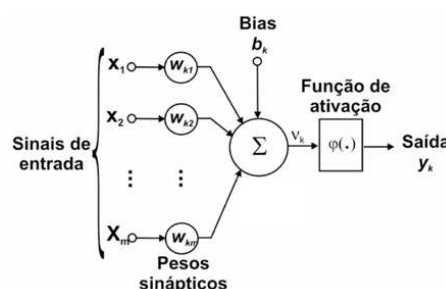


Figura 3.2: Exemplo de um neurónio artificial (Fonte: Haykin et al. (2009))

## 3.2 Algoritmos de optimização

Os algoritmos de optimização permitem a actualização de pesos das ligações dos neurónios e velocidade de aprendizagem, o objectivo destes métodos é minimizar a função de perda no caso de previsões, e maximizar a função objectivo no caso de classificação. Existem diversos algoritmos de optimização para diversos casos, e cada um tem as suas vantagens e desvantagens. Os principais algoritmos de optimização são:

- Gradiente descendente: é um algoritmo de optimização de primeira ordem que depende da derivada de primeira ordem de uma função de perda. O gradiente descendente tem como principal vantagem não ser computacionalmente exigente mas tem como desvantagens poder ficar preso num mínimo local e ao mesmo tempo, como os pesos são alterados após o

cálculo do gradiente em todo o conjunto de dados, se o conjunto de dados for muito grande, poderá levar muito tempo a convergir para o mínimo;

- Gradiente descendente estocástico (SGD): É uma variante do gradiente descendente, que tal como o nome indica, calcula o gradiente a partir de uma amostra aleatória de valores de erro ao invés do cálculo do gradiente com todos os valores de erro do conjunto de dados. No gradiente descendente estocástico os parâmetros do modelo são alterados após o cálculo da perda em cada ciclo de treino. Tal como o gradiente descendente tem como vantagem ser computacionalmente eficiente e, no seu caso, actualizar mais frequentemente os valores dos pesos. Ao mesmo tempo tem a desvantagem de poder, em alguns casos, alterar os valores da rede mesmo que o mínimo global da função de perda já tenha sido atingido;
- Momento: foi criado para reduzir a alta variação do gradiente descendente estocástico e suavizar a sua convergência. Ou seja, acelera a convergência para a direcção relevante e reduz a flutuação nas direcções irrelevantes;
- Estimador adaptativo de momento (ADAM): este método usa os momentos de primeira e segunda ordem da função de perda. Este método embora um pouco mais lento que os anteriores, tem como ideia base a diminuição da velocidade de convergência de modo a não ignorar um mínimo local e tem como desvantagem ser computacionalmente exigente.

### 3.3 Definição de hiperparâmetros

Uma fase essencial na criação de uma ANN é a definição de hiperparâmetros. Estes hiperparâmetros podem ser ajustados de três formas distintas:

- Hiperparametrização manual: consiste na alteração manual dos parâmetros da rede. Esta técnica embora seja boa para melhor compreender o funcionamento das redes neuronais, ela é muito ineficiente em termos de tempo e susceptível a preconceitos e falhas no julgamento do que é óptimo;
- Grid search: este método testa todas as combinações desejadas pelo utilizador de forma a encontrar a melhor combinação dentro dos parâmetros fornecidos, devido a ser um método exaustivo de procura não é muito apropriado para *datasets* muito grandes devido ao tempo despendido;
- Pesquisa aleatória: Simples, eventualmente são encontrados bons parâmetros. Pode ser usado para grandes conjuntos de dados, mas pode ser ineficiente pois nunca procura em todas as combinações possíveis qual a melhor para o caso em questão;
- Optimização bayesiana: usa um modelo probabilístico multivariado (representado por uma rede Bayesiana) para gerar novas soluções a cada iteração. Ao avaliar iterativamente uma configuração promissora de hiperparâmetros com base no modelo actual, e em seguida

actualizando-a, a optimização bayesiana visa obter o máximo de informação sobre a função a otimizar de modo a encontrar o seu valor óptimo;

- Algoritmo genético: usa métodos inspirados na evolução biológica, tais como a reprodução, mutação recombinação e selecção. As soluções candidatas, à solução do problema de optimização, desempenham o papel de indivíduos numa população, e a função de perda determina a qualidade das soluções. Aplicando os métodos acima referidos é assim possível gerar novas soluções até ser encontrada uma boa solução para o problema.

### 3.4 Medidas de performance

A avaliação de uma ANN é uma parte essencial de qualquer projecto de *machine learning* (ML). A avaliação é realizada através de medidas, sendo que estas estão divididas em dois tipos: medidas para classificação e medidas para previsão.

As métricas para classificação avaliam a precisão das classificações dadas pela ANN. Existem diversas métricas para este fim sendo que as principais são: (i) a precisão, que calcula o rácio entre as classificações acertadas e o total das classificações feitas; (ii) *confusion matrix* (matriz de confusão), que gera uma matriz com os verdadeiros positivos, os verdadeiros negativos, os falsos positivos e os falsos negativos; (iii) a pontuação f1 que realiza uma média harmónica entre a precisão e o *recall*. A precisão significa a percentagem dos resultados que são relevantes. Por outro lado, *recall* refere-se à percentagem do total dos resultados relevantes classificados correctamente pelo algoritmo.

As medidas para previsão avaliam a precisão das previsões da ANN. Para este tipo de avaliação existem diversos tipos de métricas sendo que as principais são: (i) o erro médio absoluto (MAE), que calcula a média da diferença entre os valores originais e os valores previstos, dando assim uma média de quão longe as previsões estão do valor real; (ii) o erro médio quadrático (MSE) que é bastante similar ao erro médio absoluto mas que por sua vez, calcula a média do quadrado da diferença entre os valores reais e os valores previstos; (iii) a raiz do erro médio quadrático (RMSE) é uma métrica quadrática que também mede a magnitude média do erro. É a raiz quadrada da média das diferenças quadráticas entre previsão e o valor real.





## Capítulo 4

# Mobilidade Urbana na Área Metropolitana do Porto

A mobilidade urbana é em si definida como a capacidade de deslocamento de pessoas com o objectivo de desenvolver actividades económicas ou sociais (Rodrigue et al. (2017)). Este conceito começou a ter maiores implicações no início dos anos 90 devido a uma crescente consciencialização das implicações, tanto históricas como contemporâneas, nas sociedades actuais, criando assim novos paradigmas nos sistemas de transporte público colectivos.

O conceito tem sofrido ao longo do tempo diversas alterações na forma como é pensado e discutido. Nos últimos anos têm sido introduzidas novas noções na análise deste conceito de mobilidade, como o estilo de vida e preferências pessoais, o que gerou uma alteração na forma como se examina a procura e a oferta de transportes públicos por parte dos governos e empresas privadas. As entidades responsáveis por gerir os transportes públicos têm procurado analisar as tendências de mobilidade de modo a melhor servirem as populações e ao mesmo tempo optimizarem as suas operações.

A Área Metropolitana do Porto (AMP) é a segunda área metropolitana mais importante de Portugal, devido à sua área e à dimensão da sua população residente.

No presente capítulo será apresentada uma caracterização geográfica e sócio-económica da Área Metropolitana do Porto (AMP) (secção 4.1), uma descrição do sistema de Transportes Públicos Intermodais do Porto (TIP) (secção 4.2) e as principais características da mobilidade da AMP (secção 4.3).

### 4.1 Área metropolitana do Porto

A Área Metropolitana do Porto (AMP) está inserida na Região Norte (NUT II), mais precisamente no Litoral Norte de Portugal e abrange uma área de aproximadamente 2 040 km<sup>2</sup> com cerca de 1 700 000 habitantes. O clima na AMP é um clima temperado com verão seco e suave, com a denominação Csb (normal climática 1961-1990). O verão nesta zona é agradável, seco e de céu quase sem nuvens, enquanto o inverno é fresco, com precipitação moderada e céu parcialmente

encoberto. Ao longo do ano, a temperatura varia entre 6 °C a 24 °C e raramente é inferior a 1 °C ou superior a 30 °C.

É composta por 17 municípios (Arouca, Espinho, Gondomar, Maia, Matosinhos, Oliveira de Azeméis, Paredes, Porto, Póvoa de Varzim, Santa Maria da Feira, Santo Tirso, São João da Madeira, Trofa, Vale de Cambra, Valongo, Vila do Conde, e Vila Nova de Gaia) e faz fronteira com as Comunidades Intermunicipais do Cávado, do Ave, do Tâmega e Sousa, de Viseu Dão Lafões e da Região de Aveiro. O município de Vila Nova de Gaia é, segundo os Censos 2011, o que apresenta maior número de residentes, detendo cerca de 17,2% do total de população. Segue-se o Porto com 13,5%, Matosinhos, Gondomar e Maia, que somam respectivamente, 10,0%, 9,5% e 7,7% dos residentes. Por outro lado, Arouca e São João da Madeira são os municípios que apresentam a menor proporção com 1,3% e 1,2%, respectivamente. Na densidade populacional, verifica-se uma grande heterogeneidade, já que os valores máximos, Porto com 5 739 hab/km<sup>2</sup>, e mínimos, Arouca 68 hab/km<sup>2</sup>, são notavelmente contrastantes.

De acordo com os dados do último recenseamento (INE (2012)), tem-se assistido ao envelhecimento da população, em que a proporção de população jovem e activa tem vindo gradualmente a diminuir, salientando-se que, em 2011, 16% dos residentes possuíam mais de 65 anos. A dimensão média das famílias decresceu em 2011, sendo que o Porto e Matosinhos registavam os valores mais baixos (respectivamente 2,32 e 2,55 elementos) e em Paredes registava-se o valor mais alto (2,97 elementos).

## 4.2 Sistema Andante

A rede TIP é suportada por um sistema de validação baseado no cartão Andante. Este cartão de validação é um *Smart Card* que serve de suporte ao sistema de bilhética intermodal usado nos transportes públicos da AMP.

O Andante é um título de transporte intermodal que permite viajar nos diversos tipos de transportes públicos da AMP, designadamente autocarro, metro, comboio e eléctrico, independentemente do trajecto que o utilizador escolhe, sendo que o valor a pagar apenas varia consoante as zonas que o utilizador pretende viajar. A subsecção 4.2.1 apresenta uma descrição do serviço prestado pela rede TIP e o seu modelo de operação enquanto que na secção 4.3 são descritas algumas características da mobilidade na AMP.

### 4.2.1 Descrição do serviço de transportes na área metropolitana do Porto

Os TIP - Transportes Intermodais do Porto é um Agrupamento Complementar de Empresas (ACE) participado pela Metro do Porto, Sociedade de Transportes Colectivos do Porto e Comboios de Portugal (CP) constituído a 20 de Dezembro de 2002 que tem como missão oferecer mobilidade dentro da AMP através de diversos meios de transporte público.

Este sistema é composto por 29 operadores, públicos e privados, que operam na AMP num total de 725 linhas. Esta rede cobre uma área aproximada de 2040 km<sup>2</sup> e é dividida em zonas que se estendem por 17 concelhos: Arouca, Espinho, Gondomar, Maia, Matosinhos, Oliveira

de Azeméis, Paredes, Porto, Póvoa de Varzim, Santa Maria de Feira, Santo Tirso, São João da Madeira, Trofa, Vale de Cambra, Valongo, Vila do Conde e Vila Nova de Gaia.

As paragens e estações de transportes públicos da AMP dos diferentes meios de transporte apresentam diferenças entre estes meios de transporte e dentro destes meios de transporte. As suas principais características diferenciadoras são a qualidade de abrigo, informação e conforto que estas fornecem aos seus utilizadores. Nas estações de autocarro e metro, Figuras (4.1 a) e 4.2 a)), é notório a existência de qualidade de abrigo contra as condições meteorológicas. Estas estações proporcionam ao utilizador conforto enquanto esperam pelo meio de transporte, nomeadamente zonas para sentar, e informação em tempo real do tempo de espera pelo meio de transporte. No caso das paragens a situação é distinta. As paragens de metro proporcionam, tal como as estações, zonas para sentar e previsões de chegada dos veículos em tempo real, sendo que o abrigo fornecido, no caso de eventos meteorológicos intensos, não é suficiente, em particular durante a hora de ponta. O mesmo se aplica ao tipo de paragem de autocarro, sendo que neste tipo não existe informação em tempo real da previsão de chegada do meio de transporte, apenas um horário informativo das possíveis horas de chegada (Figura 4.2 b) e c)). Para o tipo de paragens demonstrado na Figura 4.2 b) não existe qualquer tipo de abrigo fornecido ao utilizador nem zonas para sentar.



Figura 4.1: Exemplo de uma estação (a) e paragem (b) de metro da AMP (Flickr (2019)).



Figura 4.2: Exemplo de estação de autocarro (a) e paragens (b e c) de autocarro da AMP (Porto. (2018); JPN (2019))

Observando as diferentes estações e paragens dos dois meios de transporte nota-se que, de modo geral, as estações e paragens de metro oferecem uma maior protecção contra as condições meteorológicas adversas, e ao mesmo tempo proporcionam um maior conforto ao utilizador e, tal como dito anteriormente, fornecem sempre a hora de chegada em tempo real.

Os utilizadores podem optar pela compra de um agente único, válido apenas para uma viagem, ou por títulos mensais. Os agentes únicos podem ser adquiridos por qualquer utilizador que pretenda realizar uma viagem na rede de transportes. O valor do agente único varia consoante o número de zonas a que pretende ter acesso, sendo que o tempo de viagem é dependente do número de zonas. Para os turistas que pretendam viajar dentro da AMP, os TIP disponibilizam um agente único especial denominado "Andante Tour" que após a sua compra o utilizador pode viajar na AMP durante 24h ou 72h livremente sem limite de validações. De notar que desde 1 de Maio de 2019 o zonamento do Andante foi alterado permitindo aos utilizadores conhecerem de forma mais intuitiva os limites das zonas. Na Figura 4.3 é apresentado o mapa de zonamento do sistema de validação Andante.

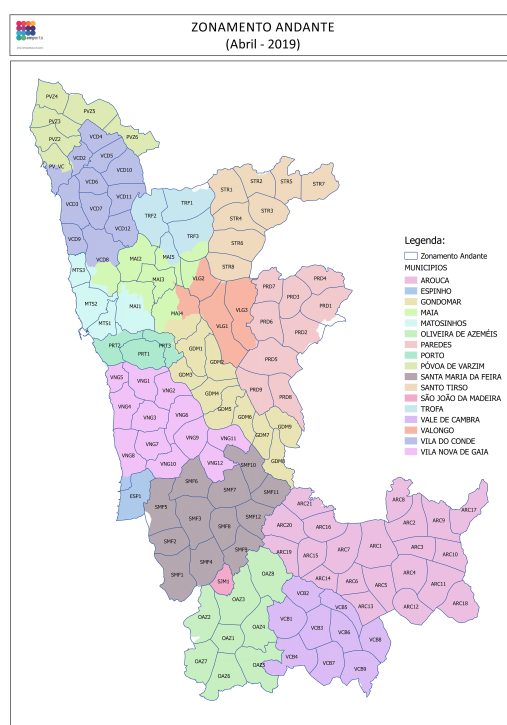


Figura 4.3: Mapa do zonamento Andante (AMP (2019))

Os títulos mensais estão divididos em: (i) monomodal e (ii) intermodal. No sistema monomodal, o utilizador pode apenas viajar nos serviços oferecidos pelo operador que escolherem. Por sua vez no sistema intermodal, o seu uso é mais alargado, pois no ato de compra podem ser escolhidas as diferentes zonas às quais o utilizador quer ter acesso independentemente do fornecedor do serviço.

Ambos os títulos mensais por sua vez estão divididos em um de dois tipos: (i) normal e (ii) social. Os títulos sociais são divididos em diversos escalões, nomeadamente:

- Júnior: Destinada a todas as crianças até aos 12 anos (inclusive);
- Estudante: Destinada a todos os estudantes até aos 25 anos (inclusive) matriculados em cursos reconhecidos oficialmente de ensino primário, secundário, técnico-profissional, médio ou superior;
- Sénior: Destinada a pessoas com idade igual ou superior a 65 anos;
- Social+: Destinada indivíduos e agregados familiares com menores recursos económicos.

Estes títulos têm atenção tanto à idade como às possibilidades económicas dos utilizadores, sendo oferecido a estes utilizadores um título mensal com valor reduzido em relação aos títulos mensais normais.

### **4.3 Características da mobilidade na área metropolitana do Porto**

Na AMP, segundo os censos realizados em 2011, a população residente nesta área demonstra uma preferência acentuada pelo uso de transporte individual em relação ao uso do transporte público (INE (2012)). De acordo com o INE (2012) este meio de transporte é utilizado, preferencialmente, por cerca de 63% dos residentes. Relativamente aos municípios que mais dependência apresentam em relação ao transporte individual, destacam-se dos demais a Maia (69,9%), Oliveira de Azeméis (72,7%), Santa Maria da Feira (72,4%) e Vale de Cambra (74,9%), cujos valores superam consideravelmente a média registada na AMP para esse ano (62,5%). Já os municípios de Espinho (55,9%), Paredes (58,9%) e Porto (51,4%) são os municípios que apresentavam uma menor proporção de utilização de transporte individual, com valores relativamente inferiores à média da AMP para 2011.

As deslocações realizadas em autocarros têm especial relevância nos concelhos de Gondomar (20%), Porto (17%) e Vila Nova de Gaia (14%), que em conjunto com os concelhos de Valongo (13%), Matosinhos (13%) e a Maia (10%), representam assim a área de influência da rede de transportes da STCP. Segundo dados dos TIP - Transportes Intermodais do Porto (2013), no que se refere à procura por zonas, constata-se que a zona C1 (Porto Centro) regista maior utilização, com 44,8% das validações totais, seguindo-se as zonas S8 (Gaia Mafamude) com 10,6%, e C2 (Porto Ocidental) com 9,7%. No seu conjunto, as 7 zonas mais centrais C1, C2, C3, C5, C6, C9 e S8 (Figura 4.4), concentram aproximadamente 88% do total de validações do Sistema Intermodal Andante (SIA). As deslocações de curta distância (clientes utilizadores de 2 ou 3 zonas) constituem a maioria das deslocações intermodais, representando 78,8% das validações efectuadas em 2013.

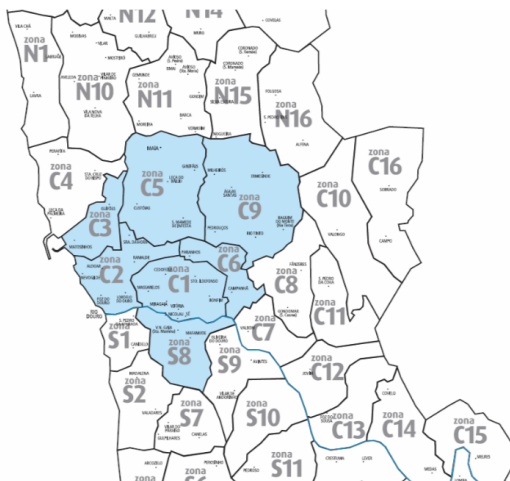


Figura 4.4: Mapa de zonas de utilização do sistema Andante na Área Metropolitana do Porto (TIP - Transportes Intermodais do Porto (2013))

## Capítulo 5

# Metodologia

As condições meteorológicas têm um grande impacto na mobilidade diária dos cidadãos. Dependendo das condições meteorológicas observadas, as viagens podem ser atrasadas e a procura e a repartição modal podem ser alteradas de forma significativa, já que em determinadas condições alguns grupos de utilizadores, tais como idosos, doentes e crianças, tendem a não viajar com a mesma frequência. Estas variações têm um grande impacto no uso e operação dos serviços de transporte público, especialmente nos serviços que funcionam próximos da sua capacidade máxima o que pode afectar a qualidade do serviço fornecido pelos operadores de transporte.

Neste capítulo descreve-se a metodologia usada para desenvolver uma ferramenta de previsão que pretende ajudar os tomadores de decisão a melhor compreender os impactos da meteorologia na procura nos transportes públicos, designadamente na procura de autocarros em zonas urbanas. Pretende-se que a ferramenta efectue a previsão da procura total na rede e da procura por linha de operação do serviço.

Para desenvolver esta ferramenta foram criados diversos cenários de modo a definir as variáveis de entrada mais robustas a usar no modelo de previsão. A secção 5.1 descreve os dados recolhidos e a secção 5.2 define os grupos de utilizadores, consoante a sua idade e apoios sociais, para os quais a ferramenta poderá efectuar previsões de procura. A secção 5.3 apresenta a metodologia para definição dos cenários enquanto a secção 5.4 apresenta o modelo, designadamente a arquitectura, métodos de aprendizagem, técnicas de optimização e medidas de performance. A Figura 5.1 apresenta o diagrama com a metodologia adoptada.

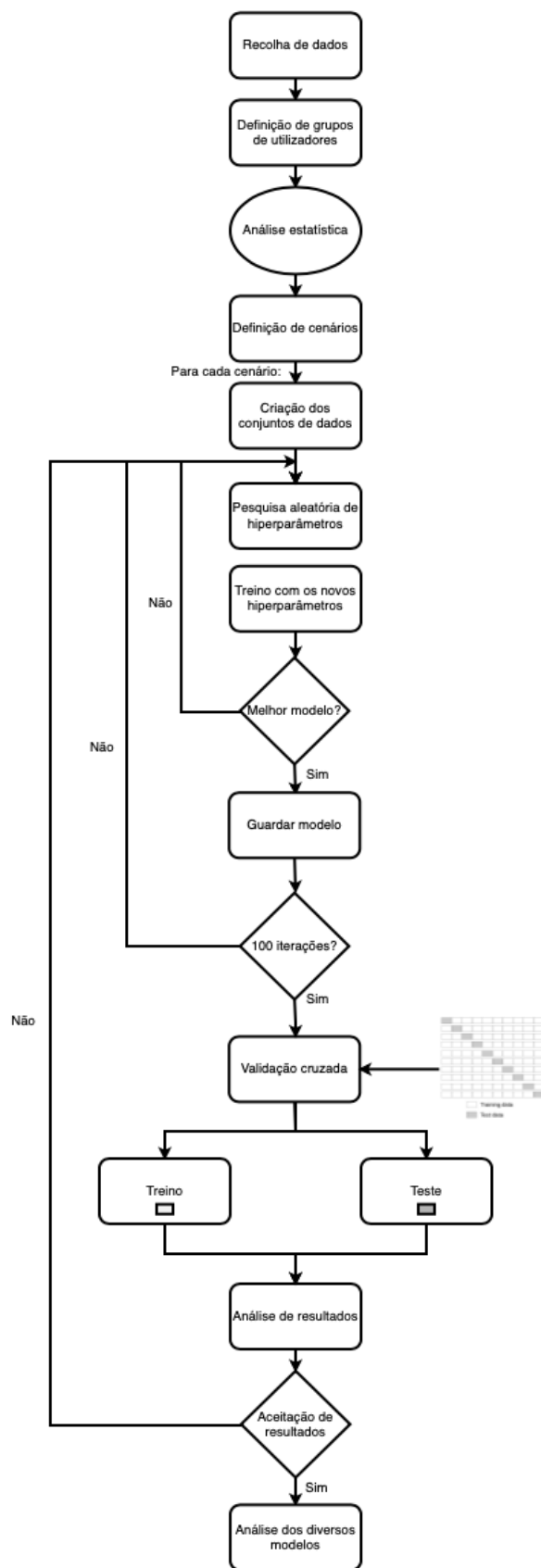


Figura 5.1: Diagrama da metodologia utilizada para definição do modelo da ANN.



## 5.1 Recolha de dados

De modo a estimar o impacto das variações meteorológicas na procura de transportes públicos foram recolhidos dados de procura de transportes públicos e condições meteorológicas. Os dados foram obtidos para a Área Metropolitana do Porto (AMP) para o período de 1 de Janeiro de 2013 a 31 de Dezembro de 2013.

Os dados da procura de transportes públicos foram obtidos a partir das validações dos títulos do Andante dos Transportes Intermodais do Porto (TIP). Os dados recolhidos (N=77.213 milhões de registos) referem-se às validações realizadas nas viagens realizadas nos autocarros da empresa STCP.

Cada registo de um cartão Andante contém informação relativa a uma viagem, nomeadamente informação do ID do cartão, da paragem e da linha, do operador do serviço, da hora da validação, do perfil do utilizador e das zonas a que o utilizador tem acesso. A Tabela 5.1 apresenta um exemplo desta estrutura de dados.

Os dados meteorológicos foram recolhidos na estação meteorológica automática n. 545 de Pedras Rubras. A estação localiza-se no centro da AMP, pelo que os seus dados são considerados representativos das condições meteorológicas na área em estudo pois encontra-se dentro dos limites de representatividade (10 km a 20 km) (Orlanski (1975)).

As variáveis meteorológicas obtidas foram: temperatura (°C), velocidade do vento (Km/h), humidade relativa (%), ponto de orvalho (°C), pressão atmosférica (mb) e nebulosidade. Os dados obtidos apresentam uma periodicidade de meia hora.

Do modo a obter um conjunto uniforme, os dados foram pré-processados e normalizados. Em primeiro lugar, os registos onde se verificou a falta de alguma variável meteorológica foram removidos. Posteriormente, os dados meteorológicos foram normalizados de modo a variarem entre 0 e 1. Estas variáveis foram normalizadas pois os limites inferiores e superiores são diferentes para cada uma das variáveis, o que pode influenciar negativamente a aprendizagem do modelo.

No final foram obtidos aproximadamente 793 mil registos que irão fazer parte dos diversos conjuntos de dados.

A Tabela 5.2 apresenta as variáveis recolhidas, nomeadamente as unidades e periodicidade.

Tabela 5.1: Exemplo da estrutura de dados de validações do sistema Andante

Operador	ID cartão	Paragem	Linha	Sentido	Variante	Veículo	Data Hora Validação	Grupo	Perfil	Zona	Zonas
STCP	20007188034	4207	907	1	1	3032	30/01/13 20:19	Título de Viagem	Normal	2	Z3 C1
STCP	20031228054	2211	907	2	1	3032	30/01/13 20:36	Assinatura	Normal	2	C1 S8-9
STCP	20031312314	2085	907	2	1	3032	30/01/13 20:37	Assinatura	Estudante	2	C1 S8-9
STCP	20031259162	2085	907	2	1	3032	30/01/13 20:37	Assinatura	Social+	2	C1 S8-9
STCP	20032152919	2184	907	2	1	3032	30/01/13 20:38	Assinatura	Social+ R	2	C1-2 S8-9
STCP	20032090190	3599	902	1	1	3215	31/01/13 21:24	Assinatura	sub23	2	C1 C6 S1 S8
STCP	20032166615	3599	902	1	1	3215	31/01/13 21:24	Assinatura	Estudante	2	C1 C6 S1 S8

Tabela 5.2: Descrição das variáveis usadas nas definições dos cenários

Variáveis	Sigla	Descrição/Unidade	Periodicidade
Procura	Procura	nº de validações	30 min
Hora	Hora	01h:00m; 01h:30m ...	30 min
Dia da semana	DS	Segunda-feira : 1; Terça-feira: 2 ...	-
Mês	Mês	Janeiro: 1; Fevereiro: 2 ...	-
Greve	Greve	Sim: 1; Não :0	-
Feriado	Feriado	Sim: 1; Não :0	-
Temperatura	T	°C	30 min
Velocidade do vento	WS	km/h	30 min
Humidade Relativa	RH	%	30 min
Ponto de Orvalho	DT	°C	30 min
Pressão Atmosférica	P	mb	30 min
Nebulosidade	CL	Limpo; Algumas Nuvens; Chuva...	30 min

## 5.2 Definição de grupos

Diferentes tipos de utilizadores tendem a reagir de forma distinta às alterações das condições meteorológicas, o que pode influenciar a sua predisposição a viajar (Hjorthol (2013), Mahrsi (2014)). Assim, com o objectivo de melhor compreender os impactos das variáveis meteorológicas nos diferentes tipos de utilizadores, estes foram segmentados em quatro grupos. Os diversos grupos foram criados tendo em conta a faixa etária, a ocupação e a existência de apoios sociais:

- G1 (estudantes): inclui estudantes com idades compreendidas entre os 4 anos e os 23 anos;
- G2 (normal): inclui trabalhadores em geral e validações realizadas com agentes únicos;
- G3 (idosos): inclui utilizadores reformados e utilizadores com mais de 65 anos;
- G4 (social): inclui utilizadores aos quais são atribuídos apoios sociais.

## 5.3 Definição de cenários para escolha das variáveis de entrada no modelo

Em situações reais é geralmente difícil recolher uma grande diversidade de variáveis, tal como no caso de estudo que aqui se apresenta. Assim, é importante compreender o impacto do modelo quando apenas um número reduzido de variáveis é usado. De modo a seleccionar as variáveis com maior influência na capacidade de previsão da procura, quatro cenários foram definidos considerando diferentes *inputs*.

Para definição dos cenários foi efectuada uma análise estatística das variáveis de entrada. Aplicou-se o teste de correlação de Spearman ( $r_s$ ) (teste não-paramétrico) uma vez que, pela aplicação do teste não-paramétrico de Kolmogorov-Smirnov, que permite comparar a distribuição de uma amostra com uma distribuição de referência (neste caso distribuição normal), se verificou que

nenhuma das variáveis em estudo segue a distribuição normal ( $p < 0.001$ ) (Tabela B.1). Esta análise foi realizada usando o software de análise estatística IBM SPSS Statistics for Windows (2019) (26.0.0.0).

A correlação de Spearman permite avaliar a importância e direcção de associação entre as diferentes variáveis, nomeadamente se a meteorologia tem influência na ocorrência de uma viagem a ser realizada de imediato, na hora seguinte ou nas duas horas seguintes. Para além das variáveis meteorológicas e da procura foram ainda analisadas quatro variáveis: dia da semana (DS), greve, feriado e mês. Estas variáveis foram consideradas pois habitualmente tendem a influenciar a variação da procura de transportes públicos.

Na Tabela 5.4 são apresentados os resultados do teste de correlação de Spearman. Através da análise da tabela verifica-se que a hora, a temperatura (T), a velocidade do vento (WS) e a humidade relativa (RH) são as variáveis com maior correlação significativa com a procura de autocarros. Por outro lado, a pressão atmosférica ( $r_s(\text{Idosos}) = 0,012$ ;  $r_s(\text{Estudantes}) = 0,026$  e  $p < 0,01$ ;  $r_s(\text{Social}) = 0,002$ ;  $r_s(\text{Normal}) = -0,003$ ) aparenta não ter uma correlação directa com a procura. No que diz respeito aos estudos onde existe um atraso de um hora ou um atraso de duas horas, verifica-se que as correlações das variáveis com a procura tendem a diminuir, à excepção da nebulosidade onde a correlação se torna muito significativa ( $r_s < 0,01$ ) nos grupos de utilizadores Social e Normal.

Com base nos resultados obtidos foram definidos quatro cenários (Tabela 5.3). No cenário C1 a procura de transportes públicos é estimada com base nas variáveis meteorológicas medidas na mesma hora que as viagens ocorrem. Nos cenários C2 e C3 as estimativas da procura são realizadas com base na hora anterior e nas duas horas anteriores, respectivamente. Estes cenários permitem simular a influência da meteorologia actual numa viagem a ser realizada na hora seguinte ou nas duas horas seguintes. No cenário C4 foram apenas usados como dados de entrada no modelo: o mês (Mês), o dia da semana (DS), a hora da validação (Hora), se foi ou não feriado (Feriado) ou greve (Greve) e a procura verificada. Este último cenário (C4) permite analisar o impacto do uso das variáveis meteorológicas na exactidão da previsão da procura de transportes públicos. A Tabela 5.3 apresenta as variáveis de entrada no modelo definidas para cada um dos cenários considerados.

Tabela 5.3: Cenários aplicados para a procura total e para a procura por linha de operação

Variáveis											
Cenários	Hora	DS	Greve	Feriado	Mês	T	WS	RH	DT	P	CL
C1	x	x	x	x	x	x	x	x	x	x	x
C2 (atraso de 1h)	x	x	x	x	x	x	x	x	x	x	x
C3 (atraso de 2h)	x	x	x	x	x	x	x	x	x	x	x
C4	x	x	x	x	x						

Nota:

Variáveis: Hora - hora da validação; DS - dia da semana; Greve - dia de greve; Feriado - dia de feriado; Mês - mês da validação; T - temperatura; WS - velocidade do vento; RH - humidade relativa; DT - ponto de orvalho; P - pressão atmosférica; CL - nebulosidade;

Tabela 5.4: Correlação de Spearman aplicado à procura da rede de transportes públicos (autocarros)

Tipo de Estudo	Variáveis	Grupo			
		Procura Idosos	Procura Estudantes	Procura Social	Procura Normal
Real	Hora	,317**	,486**	,370**	,384**
	T	,324**	,206**	,253**	,253**
	WS	,272**	,249**	,225**	,221**
	RH	-,372**	-,305**	-,276**	-,275**
	DT	,055**	-0,004	,053**	,056**
	P	0,012	,026**	0,002	-0,003
	CL	0,007	0,009	,018*	,018*
	DS	-,147**	-,168**	-,208**	-,199**
	Greve	-,066**	-,065**	-,074**	-,085**
	Feriado	-,063**	-,089**	-,092**	-,093**
	Mês	-,022**	-,026**	-0,001	-,025**
Atraso de 1h	Hora	,240**	,368**	,218**	,239**
	T	,289**	,188**	,216**	,219**
	WS	,239**	,246**	,205**	,205**
	RH	-,323**	-,285**	-,235**	-,239**
	DT	,045**	-,018*	,034**	,036**
	P	0,010	,021**	-0,004	-0,009
	CL	0,001	0,008	,019*	,018*
	DS	-,118**	-,171**	-,211**	-,204**
	Greve	-,070**	-,076**	-,081**	-,086**
	Feriado	-,057**	-,091**	-,093**	-,094**
	Mês	-,024**	-,027**	-0,003	-,027**
Atraso de 2h	Hora	,054**	,203**	,039**	,058**
	T	,228**	,162**	,176**	,183**
	WS	,192**	,230**	,173**	,177**
	RH	-,259**	-,265**	-,197**	-,206**
	DT	,020*	-,039**	0,012	0,015
	P	0,008	,020*	-0,007	-0,011
	CL	0,008	0,014	,025**	,026**
	DS	-,115**	-,168**	-,208**	-,201**
	Greve	-,070**	-,075**	-,081**	-,086**
	Feriado	-,057**	-,091**	-,093**	-,094**
	Mês	-,024**	-,028**	-0,003	-,027**

\*\*.Correlation is significant at the 0.01 level (2-tailed).

\*.Correlation is significant at the 0.05 level (2-tailed).

## 5.4 Modelação

As redes neuronais artificiais (ANN) são modelos estatísticos inspirados na biologia das redes cerebrais. Estes modelos têm sido usados com sucesso em diversas áreas do conhecimento, desde a engenharia às ciências sociais (Zupan and Gasteiger (1993), Cochocki and Unbehauen (1993), Rowland and Vrbka (2016)).

Neste trabalho usaram-se as ANN dado terem a capacidade de aprender e modelar relações não lineares complexas, o que é realmente importante porque na vida real muitas relações entre variáveis de entrada e saída são não lineares e são também complexas. Outra vantagem do uso destas redes é a sua capacidade de generalização e, ao contrario de outras técnicas de previsão, as ANN não impõem restrições às variáveis de entrada. Além disso, as ANN conseguem modelar mesmo com dados com heterocedasticidade, ou seja, dados com alta volatilidade e variância não constante, dada a sua capacidade de aprender relacionamentos ocultos nos dados sem impor nenhum relacionamento fixo nos dados.

O modelo foi implementado usando a linguagem de programação Python. Esta é uma linguagem muito versátil e com um vasto leque de bibliotecas que ajudam no desenvolvimento de ANN. Para o desenvolvimento das redes neuronais usou-se a biblioteca KERAS (Chollet (2015)). Devido à sua estrutura modular, esta biblioteca permite um desenvolvimento rápido das ANN, com um maior número de iterações e criação de modelos de forma intuitiva recorrendo a módulos que geram camadas de neurónios, funções de activação, funções de custo, esquemas de inicialização, esquemas de regularização e optimizadores.

Nesta secção é descrito o modelo usado, como é que foi optimizado e como é que foi aplicado para prever a procura do uso de autocarros públicos numa área metropolitana.

### 5.4.1 O modelo

Considerou-se uma arquitectura baseado na *Multilayer Perceptron (MLP) feedforward artificial neural network* que permite mapear uma série de variáveis de entrada e obter um conjunto de saídas através de uma composição de funções não-lineares. A arquitectura do tipo MLP pode ser vista como um grafo composto por uma série de camadas, por sua vez compostas por uma série de unidades de processamento (neurónios). As camadas entre as entradas e as saídas designam-se de camadas escondidas.

Para cada um dos cenários definidos na secção 5.3 definiu-se uma arquitectura de uma ANN baseada numa arquitectura base que pode ser representada na seguinte forma:  $d : n_{hid} : n_{hid} : n_{hid} : 4$ , em que  $d$  representa o número de variáveis de entrada (e que depende da complexidade de cada cenário definido na secção 5.3),  $n_{hid}$  é o número de neurónios de cada camada escondida e 4 neurónios de saída para estimar a procura para cada grupo de utilizador previamente definido na secção 5.2. Esta arquitectura foi usada pois segundo Badgwell et al. (2018) esta revela-se vantajosa na resolução deste tipo de problemas. As duas primeiras camadas têm como função extrair a informação mais importante, enquanto que as duas ultimas têm como função transformar as informações extraídas em saídas da rede.

As funções de activação introduzem uma componente não-linear nas redes neuronais, que lhe permite aprender mais do que relações lineares entre as variáveis dependentes e independentes. Para treinar redes neuronais profundas (DNN) usando retro-propagação dos erros é necessária uma função de activação que se assemelhe e actue como uma função linear, mas é, de facto, uma função não-linear que permita que relações complexas nos dados sejam aprendidas. Uma das funções de activação mais usadas nestes casos é a função de activação linear rectificada (ReLU). Apesar desta função linear permitir que relações complexas nos dados sejam aprendidas, esta converte os valores negativos em zeros, o que diminui a capacidade do modelo de ajustar ou treinar a partir dos dados correctamente. Esta transformação de valores afecta o resultado já que não mapeia os valores negativos adequadamente. Para solucionar esta questão usou-se como função de activação das camadas escondidas a Leaky ReLU (Figura 5.2). Como proposto por Maas and Ng (2013), foi adicionado à parte negativa do domínio da função de activação ReLU um declive ( $\alpha=0.01$ ) de modo a corrigir o problema anteriormente mencionado.

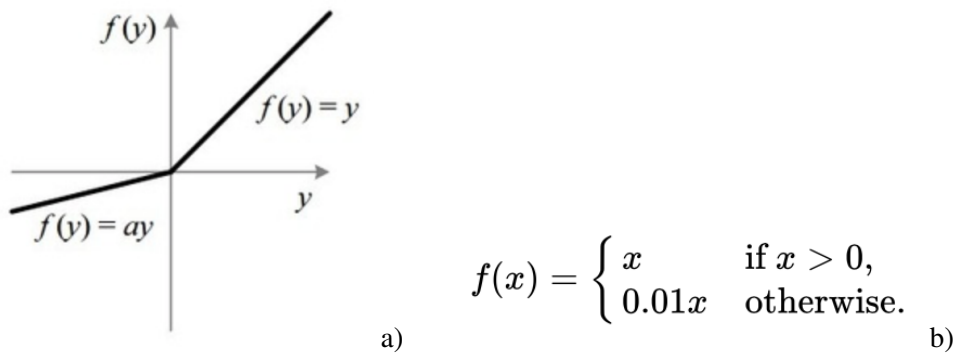


Figura 5.2: Gráfico (a) e função (b) da função de ativação Leaky ReLU.

### 5.4.2 Aprendizagem

Para a procura do conjunto de pesos óptimos (também conhecida de fase de treino) usou-se o algoritmo iterativo de retropropagação do erro (batch backpropagation) juntamente com uma técnica de optimização (subsecção 5.4.3).

A função de custo (ou função de perda) usada foi o erro quadrático médio (MSE). Esta função permite mapear um evento ou valores de uma ou mais variáveis num número real, representado assim um custo associado ao evento. A escolha da função de custo é um passo muito importante no desenvolvimento de uma ANN de previsão, pois é através dos valores resultantes desta função que se pode saber o quão próximo ou não o modelo de previsão está dos valores reais. Escolheu-se o MSE, ao invés do erro médio absoluto (MAE), pois: (i) ao elevar os erros ao quadrado antes de fazer a média, os erros mais elevados têm um peso mais elevado no calculo do erro final; (ii) um dos grandes problemas do MAE, é que o seu gradiente é o mesmo independentemente do valor do erro, o que significa que o gradiente será grande mesmo para valores de perda pequenos. Ao ser usado o MSE este problema não se aplica e conjuntamente com o algoritmo de optimização, a aprendizagem é mais rápida e eficaz.

O treino do modelo foi realizado recorrendo ao método de aprendizagem supervisionada, onde são fornecidos ao modelo valores de entrada e os valores de saída esperados, a fim de aprender com os exemplos fornecidos (Badgwell et al. (2018)). Para a previsão da procura total o modelo usou 70% do conjunto de dados para treino, 15% do conjunto de dados para validação e 15% do conjunto de dados para teste. Para a previsão da procura por linha de operação, o modelo usou 95% do conjunto de dados para treino, 2,5% do conjunto de dados para validação e 2,5% do conjunto de dados para teste. Como o conjunto de dados da procura de autocarros por linha é muito maior que o conjunto de dados da procura total de autocarros, é possível definir uma percentagem maior de dados para o conjunto de treino do modelo que prevê a procura por linha de autocarro, sem perder qualidade e quantidade de dados de treino e teste.

Na fase de treino e teste foi usado como back-end o PlaidML (0.6.4), desenvolvido pela Ver-tex.AI (2019), ao invés do TensorFlow de modo a ser possível o treino numa GPU AMD.

### 5.4.3 Optimização

O algoritmo de optimização tem como função a actualização dos pesos de modo a minimizar a função de custo. A função de custo actua como valor guia, ajudando assim o algoritmo de optimização a mover-se na direcção que permita alcançar o mínimo global. Embora o gradiente descendente estocástico (SGD) permita obter bons resultados, algoritmos de optimização mais avançados, como o ADAM (Adaptive Moment Estimation) (Kingma and Ba (2014)) ou o Adagrad (Duchi et al. (2011)), são mais rápidos a convergir numa boa solução, especialmente para o treino de ANN profundas.

Neste trabalho usou-se como algoritmo de optimização o ADAM. Este é um algoritmo com taxa de aprendizagem adaptativa que pode ser visto como uma combinação da propagação da raiz quadrática do erro médio (RMSprop) e SGD com momento. O ADAM usa os gradientes ao quadrado determinar os valores da taxa de aprendizagem, tal como RMSprop, e tira proveito do momento usando a média móvel do gradiente ao invés do próprio gradiente. O ADAM converge de forma rápida devido à sua taxa de aprendizagem dinâmica, obtendo bons resultados. Nas Figuras 5.3 e 5.4 são apresentadas as formulas de cálculo deste algoritmo de optimização.

$$\hat{m}_t = \frac{m_t}{1 - B_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - B_2^t}$$

Figura 5.3: Momento de primeiro e segundo grau

Para definir o número de épocas de treino (iterações do algoritmo de optimização), o número de neurónios das camadas escondidas ( $n_{hid}$ ) e o valor de dropout foi efectuado um conjunto de experiências preliminares. Segundo Bergstra and Bengio (2012), o uso de uma pesquisa aleatória



$$\theta_{t-1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \varepsilon} \hat{m}_t$$

Figura 5.4: Actualização de parâmetros

é mais adequado pois nem todos os hiperparâmetros são importantes para um bom funcionamento das redes neuronais.

Para definição das arquitecturas das ANN de cada um dos cenários foram seguidos os seguintes passos:

1. Pesquisa aleatória de hiperparâmetros durante 100 iterações: treino e teste do modelo usando hiperparâmetros escolhidos aleatoriamente. A cada iteração é guardado o conjunto de hiperparâmetros do modelo caso obtenha melhores resultados que os anteriores;
2. Validação cruzada: o modelo com melhores resultados é utilizado para realizar uma avaliação usando 10 fold cross-validation (validação cruzada);
3. Avaliação dos resultados da validação cruzada: caso os resultados sejam bons o modelo é escolhido para o cenário em causa, no caso contrário o modelo é descartado e é reiniciada a pesquisa aleatória de hiperparâmetros.

Na pesquisa aleatória foram procuradas as melhores configurações dos seguintes hiperparâmetros: número de neurónios por camada, número de épocas de treino, número de amostras por lote de treino (batch size) e valor de dropout entre camadas. Esta pesquisa foi realizada com gamas diferentes para cada um dos hiperparâmetros, designadamente:

- Número de neurónios: o número de neurónio de cada camada escondida variou entre 50 e 500 neurónios;
- Épocas: o número de épocas de treino variou entre 100 e 600 épocas;
- Batch size: o tamanho do batch size variou entre 8 e 100 registos para os cenários da procura total e entre 15000 e 25000 registos para os cenários da procura por linha;
- Dropout por camada: o valor de dropout em cada camada escondida variou entre 0 e 0,6.

A pesquisa aleatória foi realizada usando o algoritmo de optimização HYPERAS (0.4.1) (Pumperla (2019)). Este algoritmo avalia diversas configurações de forma aleatória, tendo em conta os parâmetros fornecidos, com o objectivo de gerar uma boa configuração dos hiperparâmetros.

#### 5.4.4 Indicadores de *performance*

As métricas de performance utilizadas foram: a raiz do erro médio quadrático (RMSE), o erro médio absoluto (MAE), o coeficiente de determinação ( $R^2$ ), e os valores médios dos erros das previsões de cada um dos grupos de utilizadores.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad ; \quad \text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$

O RMSE é um dos indicadores mais importantes em problemas de previsão. Ao contrário do MAE, os erros do RMSE são elevados ao quadrado, somados e calculada a sua raiz quadrada antes de ser calculada a média, o que faz com que o RMSE atribua um peso relativamente alto a erro elevados. Isto significa que o RMSE é bastante útil na avaliação de modelos onde erros grandes são particularmente indesejados. O coeficiente de determinação ( $R^2$ ) permite determinar se o modelo usado se ajusta bem aos dados fornecidos. O seu valor representa a percentagem de variação na resposta que é explicada pelo modelo.

Adicionalmente foi também calculado o erro médio da procura de autocarros de cada um dos quatro grupos de utilizadores da rede de transportes. Este cálculo é realizado de modo a avaliar a precisão da previsão para cada um dos grupos de utilizadores. De notar que este indicador requer que se tenham em mente que as médias das procuras de cada cada grupo de utilizadores não são iguais. Portanto, um valor menor no erro médio da procura não indica necessariamente que o modelo seja mais preciso para esse grupo em específico.

De modo a analisar os resultados das métricas descritas anteriormente é importante saber quais são os seus valores óptimos. Para o RMSE, o MAE e os erros médio por grupo de utilizador o valor objectivo é 0. Em sentido contrário, o valor objectivo do coeficiente de correlação é o valor mais alto possível, sendo que em estudos que pretendem prever comportamento humano este valor tende a situar-se em torno de 50% e para processos físicos tende a situa-se em torno de 90%.

## Capítulo 6

# Resultados

Este capítulo apresenta e discute os resultados das validações cruzadas dos modelos anteriormente definidos (secção 6.1). Posteriormente, os modelos com os melhores resultados, para cada um dos dois tipos de procura que se pretende prever, são novamente avaliados para diferentes tipos de dia (greve, feriado, semana e fim de semana) e estações do ano (secção 6.2). Estes testes têm como objectivo avaliar a robustez dos modelos a serem integrados na ferramenta final de previsão para diferentes casos, designadamente analisar a sua capacidade de previsão e generalização.

### 6.1 Avaliação da *performance* do modelo

As Tabelas 6.1 e 6.2 apresentam a arquitectura do modelo e os hiperparâmetros usados para prever a procura total e por linha respectivamente de cada cenário definido. As subsecções 6.1.1 e 6.1.2 apresentam e analisam as medidas de performance obtidas para cada tipo de previsão usando estas configurações.

Tabela 6.1: Resultados da pesquisa aleatória para previsão da procura total.

Cenário	Batch size	Épocas	Nº de neurónios na 1ª camada escondida	Nº de neurónios na 2ª camada escondida	Nº de neurónios na 3ª camada escondida	Dropout 1	Dropout 2	Dropout 3
C1	50	450	230	230	230	0,15	0,2	0,3
C2	70	350	300	300	300	0,1	0,1	0,1
C3	80	250	250	250	250	0,15	0,15	0,35
C4	60	400	150	150	150	0,126	0,12	0,11

Tabela 6.2: Resultados da pesquisa aleatória para previsão da procura por linha.

Cenário	Batch size	Épocas	Nº de neurónios na 1ª camada escondida	Nº de neurónios na 2ª camada escondida	Nº de neurónios na 3ª camada escondida	Dropout 1	Dropout 2	Dropout 3
C1	22000	300	150	150	150	0,35	0,3	0,1
C2	17000	350	300	300	300	0,35	0,2	0,2
C3	20000	500	200	200	200	0,15	0,15	0,25
C4	10000	450	100	100	100	0,2	0,1	0,1

### 6.1.1 Avaliação dos modelos de previsão da procura em toda a rede

A Tabela 6.3 apresenta os resultados da avaliação dos modelos de previsão da procura em toda a rede de autocarros.

Como pode ser observado, os resultados das diferentes medidas de performance para cada um dos modelos são bastante similares para os cenários C1, C2 e C3, com excepção do valor do RMSE que é bastante menor no cenário C3 (caso em que existe um avanço de duas horas dos valores das variáveis meteorológicas). Observando os erros médios absolutos de cada grupo de utilizador verifica-se que o cenário C3 é o que obtém melhores resultados na previsão da procura para toda a rede de autocarros. Em comparação com o modelo de previsão da procura de táxis desenvolvido por Mukai and Yoden (2012), verifica-se que o cenário C3 obteve melhores resultados, com um erro médio percentual de 7,29%, enquanto que o modelo criado por Mukai and Yoden (2012) teve um erro médio percentual de 13,59%.

A comparação dos valores médios da procura de cada um dos grupos de utilizadores (G1-G4) ao longo do ano de 2013, permite constatar que, em termos percentuais, os valores dos erros médios, quando comparados com a procura média, são: para o grupo G1 (estudantes) de 14,72%, para o grupo G2 (normal) de 5,49%, para o grupo G3 (idosos) de 6,70% e para o grupo G4 (social) de 6,30%. A discrepância entre o valor de erro percentual do grupo G1 em relação aos restantes pode-se dever, em parte, a este grupo não ter uma rotina bem definida, ao contrário dos restantes grupos que viajam mais nas horas de ponta, no caso dos G2 e G4, e entre as 10h e as 19h, de forma quase uniforme, no caso do grupo G3 (Figura A.6).

De um modo geral, os modelos que usaram as variáveis meteorológicas como variáveis de entrada (C1, C2 e C3), tiveram um desempenho melhor do que quando estas variáveis não foram usadas (C4). Os melhores resultados da previsão da procura foram obtidos quando os modelos usam como variáveis de entrada as variáveis meteorológicas registadas duas horas antes da viagem ocorrer (cenário C3). Quando estas variáveis meteorológicas são usadas, o RMSE sofre uma diminuição de 43% e o MAE uma diminuição de 24,5% (comparação do cenário C3 e C4).

Tabela 6.3: Resultados da validação cruzada para a procura total da rede de autocarros.

Cenário	R <sup>2</sup>	MAE	RMSE	Erro médio da procura de utilizadores do grupo:			
				G1 (estudantes)	G2 (normal)	G3 (idosos)	G4 (social)
C1	0,98	81,8	159,2	99	118	67	39
C2	0,97	83,2	171,7	104	111	67	39
C3	0,97	80,6	134,2	95	119	66	39
C4	0,94	106,2	235,3	127	156	94	48

### 6.1.2 Avaliação dos modelos de previsão da procura por linha

A Tabela 6.4 apresenta a previsão da procura por linha de operação de autocarros. Através da análise desta informação verifica-se que, tal como para na previsão da procura de toda a rede de autocarros, os melhores resultados para o RMSE e MAE são obtidos para o cenário C3. O melhor resultado para o  $R^2$  é obtido pelo cenário C1, sendo a sua diferença para o cenário C3 de apenas 0,01. Adicionalmente, os melhores resultados dos erros da procura para cada grupo de utilizadores é obtida para o cenário C1, sendo a sua diferença em relação ao cenário C3 de apenas um utilizador para os grupos estudantes, idosos e social. O cenário C2 é o que obtém os piores resultados, tal como verificado anteriormente na previsão da procura de autocarros em toda a rede (Tabela 6.3).

Os modelos usados para os cenários onde as variáveis meteorológicas são usadas, tiveram um desempenho melhor do que quando estas variáveis não são usadas. Comparando o cenário com o melhor resultado, com o uso das variáveis meteorológicas (cenário C3) e sem o uso das variáveis meteorológicas (cenário C4), verifica-se que, quando as variáveis meteorológicas são usadas, o RMSE sofre uma diminuição de 7,9% e o MAE uma diminuição de 15,5%. O valor do coeficiente de determinação é menor para os cenários onde as variáveis meteorológicas não são usadas. Para estes mesmos cenários o erro da procura de cada um dos grupos de utilizadores é maior quando comparados com os cenários onde são usadas as variáveis meteorológicas.

Tabela 6.4: Resultados da validação cruzada para a procura por linha de operação de autocarros

Cenário	$R^2$	MAE	RMSE	Erro médio da procura de utilizadores do grupo:			
				G1 (estudantes)	G2 (normal)	G3 (idosos)	G4 (social)
C1	0,70	9,7	21,1	6	19	7	5
C2	0,59	9,3	19,1	7	21	8	6
C3	0,69	9,1	18,1	7	19	8	6
C4	0,62	10,8	19,7	8	21	9	6

## 6.2 Resultados dos testes para as melhores redes neuronais

Para avaliar a robustez do modelo foram criados três testes que permitem avaliar a performance das redes em diferentes tipos: estações do ano (secção 6.2.1), dias da semana (secção 6.2.2), e dias com características peculiares, tais como dias de feriado e dias de greve nos transportes públicos (secção 6.2.3). Na realização dos testes acima referidos, à exceção do teste para diferentes tipos de dias (secção 6.2.3), foram usadas 200 entradas de dados, escolhidas aleatoriamente, dos conjuntos de dados finais referidos na secção 5.1, dentro dos critérios específicos de cada um dos quatro testes. Para o teste realizado com diferentes tipos de dias, foram usados dados completos da procura ao longo do dia escolhido, de forma aleatória.

Para a procura total e por linha da rede de autocarros urbanos foram usada as ANN referentes ao cenário com melhores resultados de performance (em ambos os casos, cenário C3).

### 6.2.1 Avaliação da *performance* por estação do ano

A Tabela 6.5 apresenta os resultados dos modelos de previsão para as diferentes estações do ano.

Como pode ser observado, é no Verão que os modelos obtêm uma melhor previsão, para a procura por linha de autocarro (MAE=8,7 e RMSE=15,7) e para a rede total de operação dos autocarros (MAE=63,4 e RMSE=111,3). Tais resultados podem dever-se ao facto de que durante o verão, o grupo estudantes (G1) viajar menos e o grupo de idosos (G3) adquire nesta fase do ano padrões de mobilidade mais uniformes.

A previsão da procura por linha, é menos exacta para os meses de Outono (MAE=10,2 RMSE=18,2) enquanto que no caso da procura em toda a rede de autocarros, a previsão é menos exacta durante os meses de Primavera (MAE=72,6 e RMSE=138,1).

Tabela 6.5: Resultados da previsão da procura em diferentes estações do ano.

Estação do ano	MAE Procura total	MAE Procura linha	RMSE Procura total	RMSE Procura linha	Desv. Padrão Procura total	Desv. Padrão Procura linha
Primavera	72,6	9,2	138,1	17,8	137,5	17,8
Verão	63,4	8,7	111,3	15,7	111,2	15,6
Outono	68,1	10,2	121,6	18,2	121,6	18,1
Inverno	74,2	9,6	137,9	17,6	137,6	17,5

### 6.2.2 Avaliação da *performance* por dia da semana

Após a realização do teste para a previsão em diferentes estações do ano, foi realizado um teste de forma a avaliar quais os dias da semana onde estas redes neuronais conseguem melhores resultados. De modo a aferir sobre os resultados deste teste é importante realizar uma separação entre dias da semana e fim de semana, pois como verificado no capítulo 2, existe uma clara diminuição na procura de transportes públicos aos fins de semana. A Tabela 6.6 apresenta os resultados da avaliação da performance (MAE e RMSE) do modelo para os diferentes dias da semana.

A análise destes resultados permite verificar que, para a procura total, o dia da semana onde a previsão é mais robusta é a segunda-feira (MAE=75,5 e RMSE=137,3) e ao fim de semana é ao sábado (MAE=47,4 e RMSE=70,0). Foi considerado Sábado como o dia com melhores resultados ao fim de semana pois, embora os seus MAE e RMSE sejam maiores em comparação com o Domingo, quando comparados os respectivos MAE com a procura média de cada um desses dias o erro médio é menor ao Sábado. Comparando este modelo de previsão com o modelo criado por Mukai and Yoden (2012) verifica-se que o erro médio percentual das previsões do modelo criado (C3) tem um erro muito menor. Para os dias da semana (segunda-feira a sexta-feira) o erro médio percentual é de 15,68% para o modelo criado por Mukai and Yoden (2012) e 6% para o modelo C3. Para o fim de semana, ambos os modelos têm erros maiores ao domingo, sendo o erro médio

percentual de 19,8% para o modelo criado por Mukai and Yoden (2012) e de 10% para o modelo C3, criado nesta dissertação.

Em relação à previsão da procura por linha de autocarro, verifica-se que o dia da semana onde a previsão é mais próxima do valor real é a segunda-feira (MAE=10,6 e RMSE=18,8) e ao fim de semana é o sábado (MAE=6,9 e RMSE=11,3). O Sábado foi considerado o dia de fim de semana em que o modelo obtém melhores resultados pelos mesmos motivos apresentados para o modelo que executa a previsão da procura total da rede de autocarros.

Para os dias de fim de semana a previsão do modelo obtém piores resultados ao domingo. Tal fenómeno pode ser explicado pelo comportamento mais errático dos utilizadores pois as viagens realizadas aos fins de semana tendem a ser para lazer, não existindo assim uma rotina bem definida que possa ser afectada pelas condições meteorológicas. Durante os dias da semana, no caso da procura total, os piores resultados na previsão são obtidos à quarta-feira (MAE=80,6 e RMSE=154,1) e no caso da procura por linha de autocarro são obtidos à quinta-feira (MAE=11,2 e RMSE=19,8) e Sexta-feira (MAE=11,2 e RMSE=19,8).

Tabela 6.6: Resultados da previsão da procura de autocarros em diferentes dias da semana.

Dia da semana	MAE Procura total	MAE Procura linha	RMSE Procura total	RMSE Procura linha	Desvios. Padrão Procura total	Desvios. Padrão Procura linha
Segunda-feira	75,5	10,6	137,3	18,8	136,7	18,8
Terça-feira	78,1	11,0	137,8	19,4	137,4	19,4
Quarta-feira	80,6	11,0	154,1	19,5	154,0	19,5
Quinta-feira	84,1	11,2	153,0	19,8	152,9	19,8
Sexta-feira	78,3	11,2	138,7	19,8	138,7	19,7
Sábado	47,4	6,9	70,0	11,3	70,0	11,3
Domingo	41,8	4,9	64,7	7,9	64,2	7,9

### 6.2.3 Avaliação da procura para diferentes tipos de dias

A Tabela 6.7 apresenta os resultados da procura total e da procura por linha em diferentes tipos de dias, feriados e/ou greves. Foi realizada a previsão para: (i) um dia do ano de 2013 aleatório, (ii) um dia de fim de semana aleatório, (iii) um dia em que se verificou uma greve e (iv) um dia onde se verificava ser feriado. A linha escolhida foi a linha de autocarro 200. A escolha recaiu sobre esta linha, pois esteve em funcionamento em todos os tipos de dias em que se requereu uma previsão da procura e durante a maior parte do dia.

O dia da semana, escolhido de forma aleatória, ocorreu numa Terça-feira em que a nebulosidade era de céu limpo, com temperatura média de 21°C e uma velocidade média do vento de 8 km/h. Para este dia a previsão da procura total teve um RMSE de 121,0 e um MAE de 78,9 e um RMSE de 14,5 e um MAE de 10,0 para a procura na linha 200. Através da análise dos gráficos de dispersão contidos nas Figuras C.1 e C.2, verifica-se que, para a procura total na rede de autocarros a previsão é bastante assertiva, embora no caso da procura total para o grupo estudantes a precisão da previsão seja inferior. No caso da procura na linha 200 esta previsão teve resultados bons para os grupos normal e idosos (erro médio percentual de 11,7% e 12% respectivamente) e

resultados inferiores para os grupos estudantes e social (erro médio percentual de 28% e 17,5% respectivamente).

O dia de fim de semana, escolhido de forma aleatória, ocorreu a um sábado em que o nível de nebulosidade era de chuva, com uma temperatura média de 20°C e uma velocidade média do vento de 26 km/h. Para este dia a previsão da procura total teve um RMSE de 66,2 e um MAE de 44,9 e para a procura na linha 200 um RMSE de 8,2 e MAE de 5,7. Analisando os gráficos de dispersão das Figuras C.3 e C.4 verifica-se, mais uma vez que a previsão da procura total é bastante boa nos grupos normal, idosos e social, sendo que o erro na previsão aumenta no caso do grupo estudantes. A previsão para a linha 200 obteve resultados fracos para quase todos os grupos à exceção do grupo idosos em que o erro foi menor (erro médio percentual de 17%).

O dia, escolhido de forma aleatória, em que se verificou ser feriado ocorreu a uma quarta-feira. Neste dia o nível de nebulosidade era de algumas nuvens, com uma temperatura média de 12°C e uma velocidade média do vento de 15 km/h. Para este dia a previsão da procura total teve um RMSE de 91,7 e um MAE de 55,8 e para a procura na linha 200 um RMSE de 7,3 e MAE de 5,4. Analisando os gráficos de dispersão das Figuras C.5 e C.6 verifica-se, mais uma vez que a previsão da procura total é bastante boa nos grupos normal (erro médio percentual de 9,6%), idosos (erro médio percentual de 12,7%) e social (erro médio percentual de 10,4%), sendo menos boa para o grupo dos estudantes (erro médio percentual de 16,5%). A previsão para a linha 200 teve resultados aceitáveis em todos os grupos de utilizadores.

Por último, o dia, escolhido de forma aleatória, em que se verificou uma greve nos transportes públicos ocorreu a uma Terça-feira. Neste dia o céu esteve nublado, com uma temperatura média do ar de 15°C e uma velocidade média do vento de 11 km/h. Para este dia, os erros das previsões foram muito maiores aos apresentados até ao momento, sendo o RMSE de 284,2 e o MAE de 164,8 para a procura total e o RMSE de 17,16 e o MAE de 12,4 para a procura na linha 200. Estes resultados podem dever-se a apenas terem existido quatro dias de greve durante todo o ano de 2013, não podendo assim os modelos aprender tão bem estes casos. Os gráficos de dispersão para este dia encontram-se nas Figuras C.7 e C.8.

Tabela 6.7: Resultados da previsão para diferentes tipos de dias.

Dia	Tipo de dia	MAE Procura total	MAE Procura linha	RMSE Procura total	RMSE Procura linha	Desv. Padrão Procura total	Desv. Padrão Procura linha
09/07/13	Terça-feira - sem feriado/greve	78,9	10,0	121,0	14,5	107,9	13
28/09/13	Sábado - sem feriado/greve	44,9	5,7	66,2	8,2	66,0	7,7
01/05/13	Quarta-feira - feriado	55,8	5,4	91,7	7,3	87,2	6,4
05/03/13	Terça-feira - greve	164,8	12,4	284,2	17,2	283,4	15,6



## Capítulo 7

# Ferramenta de previsão

Para o desenvolvimento da ferramenta de previsão da procura nos autocarros públicos numa área metropolitana usando dados meteorológicos, adoptou-se a metodologia apresentada no capítulo 5. Esta metodologia permitiu identificar a melhor arquitectura e hiperparâmetros de uma ANN baseada na Multilayer Perceptron (MLP).

Assim, a arquitectura da ANN implementada, para a procura total na rede de autocarros, nesta ferramenta tem a seguinte configuração: 11 : 250 : 250 : 250 : 4. As entradas da rede são: Hora, DS, Greve, Feriado, Mês, T, WS, RH, DT, P, CL. Para a procura por linha a ANN implementada nesta ferramenta tem a seguinte configuração: 12 : 200 : 200 : 200 : 4. As entradas da rede são: Hora, ID Linha, DS, Greve, Feriado, Mês, T, WS, RH, DT, P, CL. Estas arquitecturas foram adoptadas com base na análise de uma série de cenários que indicou que a procura de autocarros urbanos consegue ser melhor prevista usando dados meteorológicos correspondentes a duas horas antes da realização de uma viagem (cenário C3).

Para obtenção dos dados de entrada na ferramenta, foi usada uma API (Interface de Programação de Aplicação) gratuita, disponibilizada pela Weatherbit.io (2020), que permite obter a previsão da meteorologia para as horas seguintes. Embora a API forneça a informação necessária à execução da ferramenta, esta API apenas permite a recepção da previsão das variáveis meteorológicas até ao final do dia em que se executa o pedido de informação.

As previsões meteorológicas são obtidas através de um ficheiro JSON (JavaScript Object Notation). A ferramenta desenvolvida lê e separa os diferentes valores das variáveis meteorológicas que darão entrada na ANN implementada. Os valores recolhidos são pré-processados de modo a converter os valores para as unidades usadas pela ferramenta, designadamente os valores da nebulosidade. Estes valores, obtidos pela API Weatherbit.io (2020) são um código numérico, compreendido entre 200 e 900, que posteriormente são convertidos pela ferramenta em valores entre 0 e 5.

Para utilizar a ferramenta, o utilizador deve inserir o dia da semana, se é greve e se é feriado de modo à ferramenta puder efectuar a previsão (Figura 7.1). Após a inserção destes dados a ferramenta prevê a procura de autocarros urbanos utilizando o modelo acima descrito para as horas correspondentes às variáveis meteorológicas recebidas da API. Após a realização da previsão

da procura total e da procura por linha de operação do serviço de transportes, as previsões são guardadas em ficheiros Excel distintos (Figura 7.2). O código da ferramenta pode ser consultado no Anexo D.

```
Using plaidml.keras.backend backend.
Qual o dia da semana (Segunda:1 , Terça:2, Quarta:3, Quinta:4, Sexta:5, Sabado:6, Domingo:7)? : 2
É dia de greve (Sim:1, Nao:0)? : 0
É dia de feriado (Sim:1, Nao:0)? : 0
INFO:plaidml:Opening device "opencl_amd_gfx804.0"
Previsao para as proximas 6 horas executada.
Por favor consulte os ficheiros gerados com os nomes PrevisaTotal.xlsx e PrevisaoPorLinha.xlsx
```

Figura 7.1: Interface da ferramenta

Na Figura 7.2 é apresentado o diagrama de funcionamento da ferramenta.

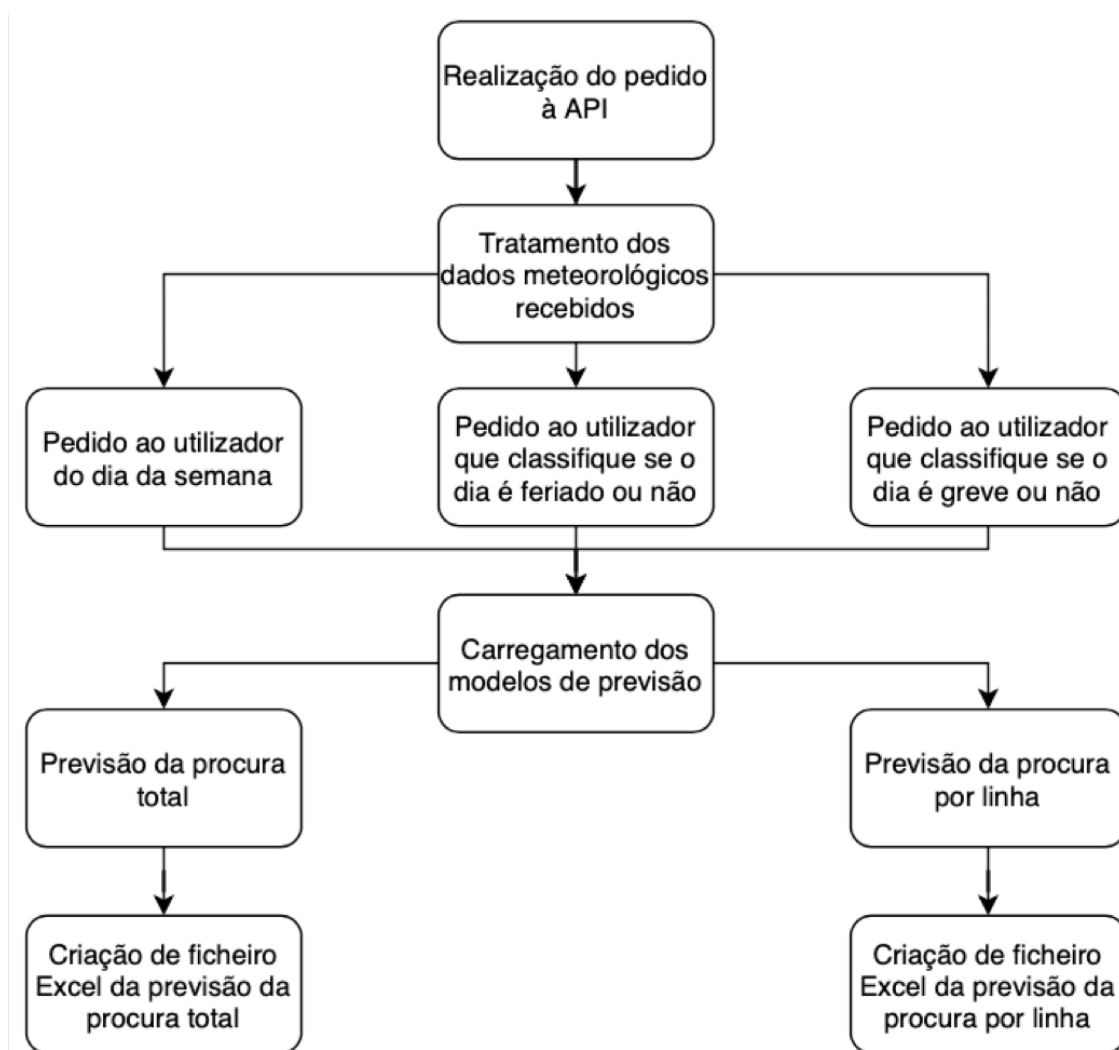


Figura 7.2: Diagrama do funcionamento da ferramenta.

## Capítulo 8

# Conclusões

Ao longo desta dissertação foram sido apresentadas algumas conclusões parciais sobre os diversos resultados. Assim, este capítulo tem como objectivo expor um sumário global das conclusões retiradas, dando especial relevo aos aspectos mais significativos, apontando-se a satisfação geral dos objectivos fixados, bem como as perspectivas de trabalho futuro.

De modo a auxiliar os tomadores de decisão a gerir o serviço de transportes de autocarros de uma área urbana tendo em conta a influência de factores meteorológicos, foi desenvolvida uma ferramenta para efectuar a previsão da procura de autocarros em toda a rede e a procura por linha de operação de autocarro.

Neste sentido foi desenvolvida uma metodologia para a obtenção dos melhores modelos de previsão, recorrendo a redes neuronais, que fazem parte da ferramenta desenvolvida. Esta metodologia começou por uma análise dos dados recolhidos, sendo posteriormente criados grupos de utilizadores para os quais a ferramenta teria de prever a procura. Posteriormente foram realizados testes estatísticos de modo a verificar quais as variáveis meteorológicas, e não só, que maior correlação tinham com a procura de autocarros, sendo no final definidos doze cenários para os quais foram criadas doze redes neuronais, uma para cada cenário. Após esta fase foi realizada a modelação dos diversos modelos, sendo procurados os melhores valores para as suas configurações e apresentados os indicadores de performance que os iriam avaliar.

Através da análise de resultados do modelo para os diversos cenários concluiu-se que o uso de variáveis meteorológicas permite aumentar a capacidade de previsão do modelo sem variáveis meteorológicas de forma substancial. Ao mesmo tempo, verifica-se que, quando aplicado um atraso de duas horas na procura em relação às variáveis meteorológicas, os resultados das previsões, para cada um dos dois tipos de procura, melhora consideravelmente, existindo uma diminuição do RMSE de 15,7% para a procura total e de 14,2% para a procura por linha. A previsão da procura total da rede de autocarros, para o melhor cenário (C3) tem um erro médio percentual, quando comparado com a procura total média, de 7,29%. A previsão da procura por linha de autocarro, para o melhor cenário (C3) tem um erro médio percentual, quando comparado com a procura média por linha, de 36,4%. Além disso, quando o modelo usa dados meteorológicos para estimar a procura com um atraso de duas horas, são obtidos melhores resultados para o coeficiente de de-

terminação, do que com os modelos em que não existe atraso ou o atraso é de apenas uma hora, o que demonstra que se ajustam melhor aos dados facultados.

Embora os resultados das previsões de um modo geral tenham sido bons (RMSE (procura total)=134,2 e RMSE (procura por linha)=18,1), verifica-se que, na previsão em ambos os tipos de procura (total e por linha), o grupo estudantes é o grupo onde as previsões não são tão assertivas. Esta situação pode ser explicada tendo em conta que este grupo não tem uma rotina bem definida, e não é tão susceptível a alterações nos padrões de viagem devido a condições meteorológicas. Verifica-se ainda que a previsão da procura é pouco exacta em dias em que se verifica greve nos transportes públicos. Apesar das ANN serem capazes de generalizar e prever para casos novos, em dias de greve onde se observa um comportamento da procura muito distinto dos dias normais e devido ao ano em estudo terem apenas existido quatro eventos destes, influencia aqui a obtenção de previsões menos exactas.

Os modelos obtidos com melhor performance foram integrados numa ferramenta que permite ajudar os decisores a compreender melhor os impactos causados pelas condições meteorológicas na procura de autocarros em toda a rede e por linha. Através do uso da API escolhida, a ferramenta obtém as previsões das condições meteorológicas para as horas seguintes, requerendo pequenas informações adicionais, tais como o dia da semana e se é dia de greve ou feriado, por parte do utilizador para o seu correto funcionamento. No final a ferramenta gera dois ficheiros Excel contendo as previsões da procura total da rede e por linha de autocarros.

Os objectivos definidos no Capítulo 1 foram plenamente cumpridos, tendo sido desenvolvida uma ferramenta de previsão da procura nos transportes públicos tendo em conta as condições meteorológicas com bons resultados e analisado o impacto na exactidão da previsão da procura com a inclusão das variáveis meteorológicas nos dados fornecidos às redes neuronais.

Um resumo do artigo sobre esta tese foi submetido à 5ª Conferência sobre Mobilidade Urbana Sustentável – CSUM2020 (<http://csum.civ.uth.gr>). O principal objectivo desta conferência é a dispersão de conhecimento e o intercâmbio de boas práticas entre investigadores e profissionais no domínio do transporte urbano. A conferência tem como principais propósitos os seguintes pontos:

- Transporte público e sistemas ajustáveis à procura;
- Remodelação de modelos de transporte;
- Comportamento dos utilizadores de transportes;
- Cidades inteligentes

Os resultados da conferência serão publicados na série de livros “Advances in Intelligent Systems and Computing” da Springer, indexada pela ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar e Springerlink. Os artigos seleccionados serão revistos e as versões estendidas serão consideradas para publicação numa edição especial.

## 8.1 Trabalho futuro

Como trabalhos futuros sugere-se o seguinte:

- Realização, através da mesma metodologia desenvolvida, da previsão da procura total e procura por linha do Metro;
- Repetir a metodologia desenvolvida para dados meteorológicos e de procura nos transportes públicos mais recentes, de modo a perceber se dados mais recentes obtêm melhores resultados;
- Avaliar a robustez do modelo para outras cidades.

Sugere-se ainda aprofundar o conhecimento sobre impacto da meteorologia nas decisões dos utilizadores de transportes públicos, através da análise e comparação da procura do Metro e dos autocarros em função das condições meteorológicas, de modo a identificar quais as condições meteorológicas que levam os utilizadores a trocar de meio de transporte público.



## Anexo A

### Análise dos dados recolhidos

No ano de 2013 foram realizadas um total de 77,2 milhões de validações, com uma média de 6,4 milhões de validações por mês e 211 mil validações por dia. A Figura A.1 apresenta o número de validações mensais realizadas em 2013 no sistema de transportes públicos andante (STCP). A análise desta figura permite verificar que existe uma clara diminuição da procura na estação de inverno e na estação de verão, sendo esta diminuição mais acentuada nos meses de Dezembro e Agosto.

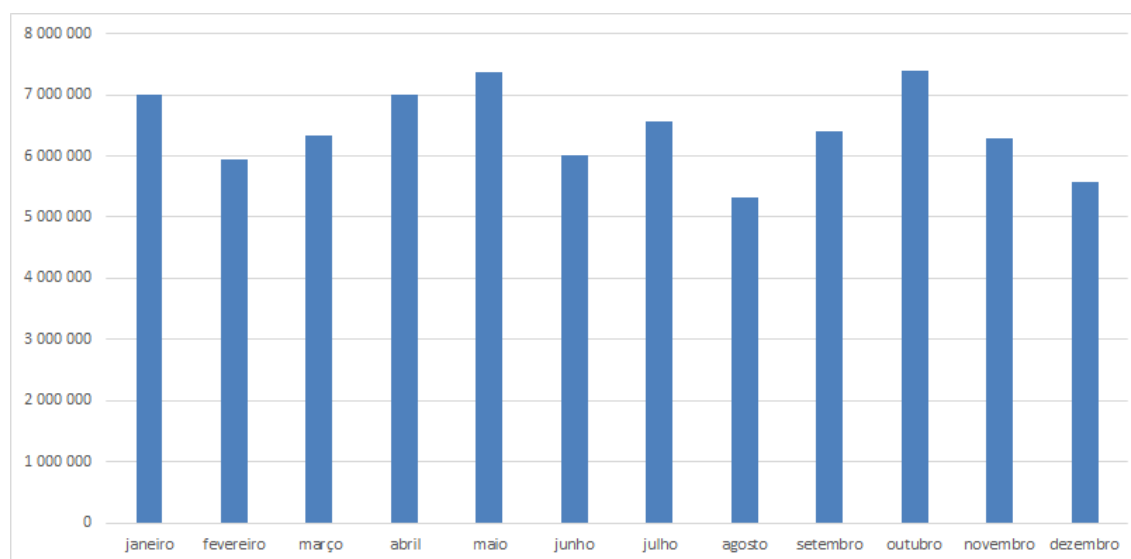


Figura A.1: Número de validações mensal no sistema de transportes públicos andante (STCP) em 2013.

Em relação ao tipo de assinatura de utilizador, existe uma clara discrepância entre os passes mensais e os títulos ocasionais, sendo que os segundos representam a maioria das validações efectuadas em todos os meses do ano de 2013. Os passes mensais registam nos meses de Julho e Agosto uma diminuição acentuada nos passes escolares, período esse que coincide com as férias escolares dos estudantes.

Para analisar as variações meteorológicas usaram-se os valores de referência da normal climatológica da região. Segundo a Organização Meteorológica Mundial (OMM), as normais climatológicas correspondem a valores médios das variáveis meteorológicas durante um período mínimo de 30 anos, de forma a obter um valor predominante que represente os padrões característicos de um parâmetro do clima no local considerado (IPMA (2019)). Estes valores médios correspondem a intervalos de tempos diários, mensais, sazonais ou anuais das variáveis meteorológicas. Segundo o Instituto Português do Mar e da Atmosfera (IPMA), as variáveis meteorológicas descrevem o comportamento físico da atmosfera, enquanto que as normais climatológicas apresentam uma informação do comportamento médio, utilizando dados de longo período para definir o clima em uma determinada época ou era em vários anos.

Comparando os valores médios mensais das variáveis meteorológicas temperatura, velocidade do vento, humidade relativa e pressão atmosférica com a normal climatológica dos anos 1961 a 1990 é possível averiguar que, em termos médios, os valores das variáveis meteorológicas não têm sofrido grandes alterações. Tais evidências podem ser observadas nos gráficos seguintes (Figura A.2, Figura A.3, Figura A.4 e Figura A.5).

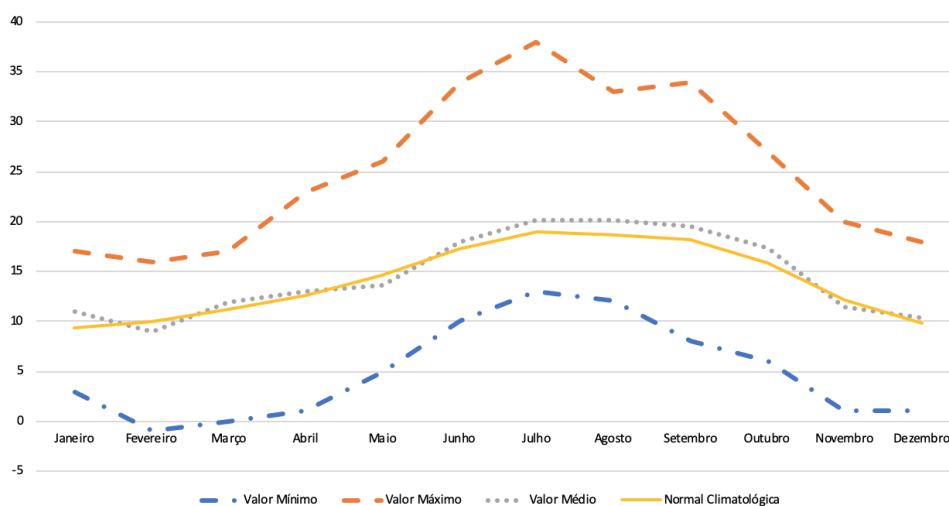


Figura A.2: Valores de temperatura (°C) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990



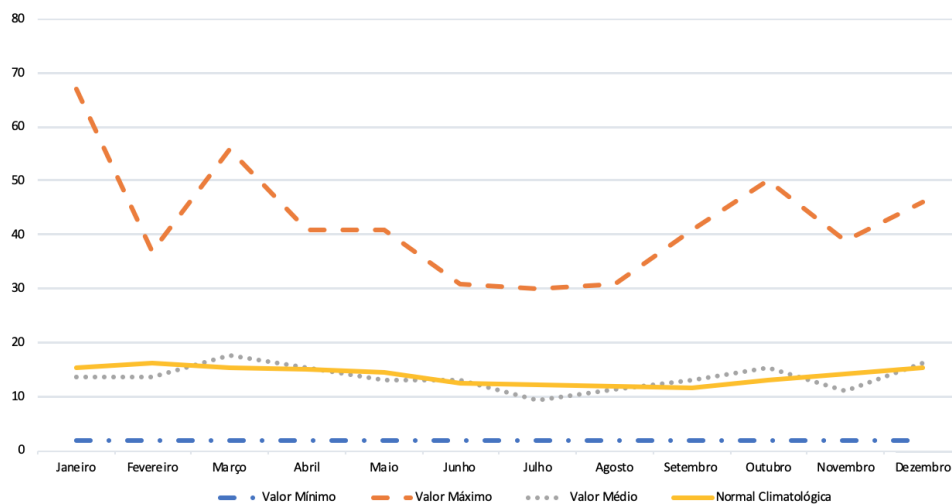


Figura A.3: Valores de velocidade do vento (Km/h) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990

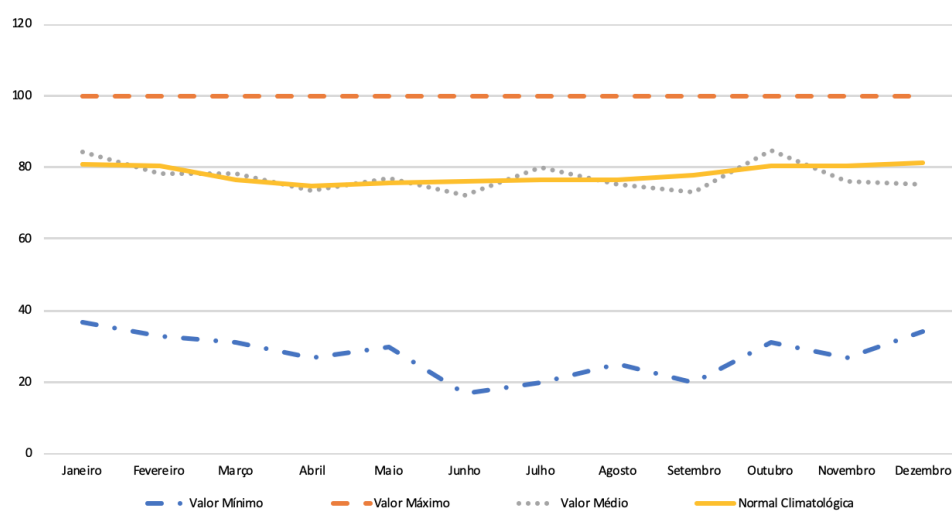


Figura A.4: Percentagem de humidade relativa (%) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990

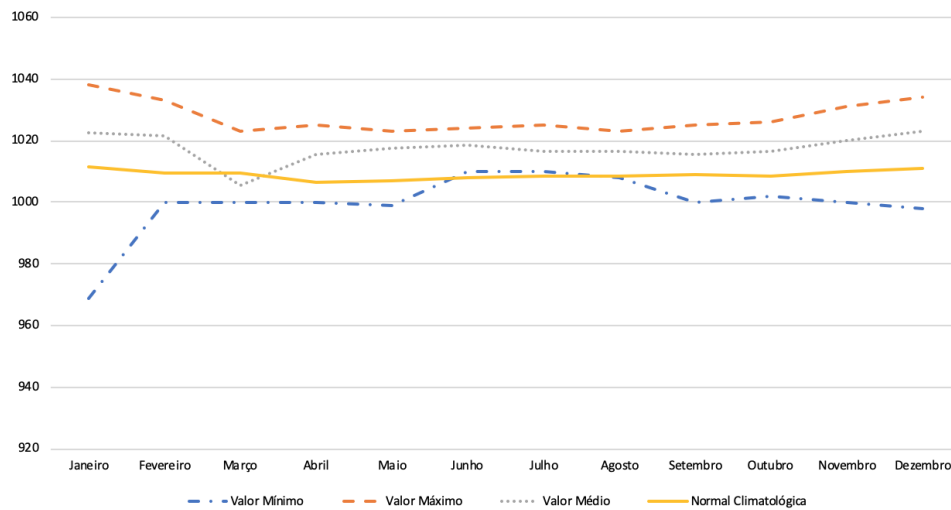


Figura A.5: Valores de pressão atmosférica (mb) ao longo do ano de 2013 e valores médios da Norma Climatológica 1961-1990

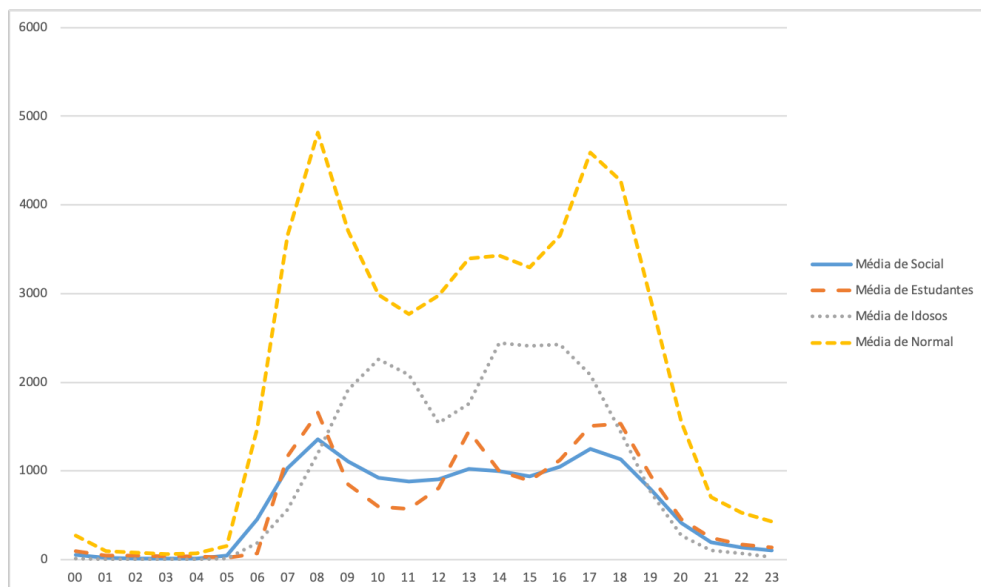


Figura A.6: Procura média por grupo de utilizador e hora durante o ano de 2013

Tabela A.1: Valores estatísticos das variáveis meteorológicas

Variável Meteorológica	Mês	Valores			Normais Climatológicas 1961 - 1990
		Mínimo	Máximo	Média	Média
Temperatura(°C)	Janeiro	3,000	17,000	10,988	9,400
	Fevereiro	-1,000	16,000	8,943	10,000
	Março	0,000	17,000	11,888	11,200
	Abril	1,000	23,000	13,071	12,600
	Maiο	5,000	26,000	13,623	14,600
	Junho	10,000	34,000	18,023	17,300
	Julho	13,000	38,000	20,164	19,000
	Agosto	12,000	33,000	20,194	18,700
	Setembro	8,000	34,000	19,576	18,200
	Outubro	6,000	27,000	17,376	15,800
	Novembro	1,000	20,000	11,478	12,100
	Dezembro	1,000	18,000	10,403	9,900
Velocidade do Vento (km/h)	Janeiro	2,000	67,000	13,726	15,400
	Fevereiro	2,000	37,000	13,633	16,200
	Março	2,000	56,000	17,592	15,400
	Abril	2,000	41,000	15,521	15,000
	Maiο	2,000	41,000	13,168	14,600
	Junho	2,000	31,000	13,168	12,400
	Julho	2,000	30,000	9,372	12,200
	Agosto	2,000	31,000	11,385	12,000
	Setembro	2,000	41,000	13,154	11,800
	Outubro	2,000	50,000	15,372	13,100
	Novembro	2,000	39,000	11,186	14,200
	Dezembro	2,000	46,000	16,246	15,500
Humidade Relativa (%)	Janeiro	37,000	100,000	84,446	81,000
	Fevereiro	33,000	100,000	78,504	80,600
	Março	31,000	100,000	78,104	76,600
	Abril	27,000	100,000	73,369	75,000
	Maiο	30,000	100,000	76,947	75,700
	Junho	17,000	100,000	72,225	76,000
	Julho	20,000	100,000	79,893	76,600
	Agosto	25,000	100,000	75,322	76,700
	Setembro	20,000	100,000	73,21	78,000
	Outubro	31,000	100,000	84,661	80,300
	Novembro	27,000	100,000	76,028	80,300
	Dezembro	34,000	100,000	75,465	81,300
Pressao Atmosferica (mb)	Janeiro	969,000	1038,000	1022,5	1011,300
	Fevereiro	1000,000	1033,000	1021,5	1009,500
	Março	1000,000	1023,000	1005,7	1009,700
	Abril	1000,000	1025,000	1015,4	1006,400
	Maiο	999,000	1023,000	1017,4	1007,000
	Junho	1010,000	1024,000	1018,4	1008,100
	Julho	1010,000	1025,000	1016,7	1008,700
	Agosto	1008,000	1023,000	1016,7	1008,500
	Setembro	1000,000	1025,000	1015,3	1009,000
	Outubro	1002,000	1026,000	1016,3	1008,700
	Novembro	1000,000	1031,000	1020,3	1010,000
	Dezembro	998,000	1034,000	1023,2	1010,800



## **Anexo B**

### **Resultados do teste Kolmogorov-Smirnov**

Tabela B.1: Teste de Kolmogorov-Smirnov

Kolmogorov-Smirnov Test												
	Hora	Procura	T	WS	RH	DT	P	CL	DS	Greve	Feriado	Mês
N	68961	68961	62124	61090	62108	62124	62124	62124	62124	62124	62124	62124
Normal Parameters a,b	Mean	11:52:42,56	15,04	13,69	76,93	10,51	1017,08	1,15	4,00	0,01	0,04	6,88
	Std. Deviation	6:53:36,203	5,437	7,730	17,607	5,125	17,771	1,272	2,008	0,102	0,194	3,277
Most Extreme Differences	Absolute	0,067	0,068	0,109	0,125	0,086	0,227	0,237	0,130	0,530	0,541	0,098
	Positive	0,063	0,068	0,109	0,095	0,063	0,226	0,237	0,128	0,530	0,541	0,096
	Negative	-0,067	-0,045	-0,069	-0,125	-0,086	-0,227	-0,184	-0,130	-0,459	-0,420	-0,098
Test Statistic	0,067	0,214	0,068	0,109	0,125	0,086	0,227	0,237	0,130	0,530	0,541	0,098
Asymp. Sig. (2-tailed)	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c	,000 c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

## Anexo C

# Resultados gráficos da avaliação da procura para diferentes tipos de dias

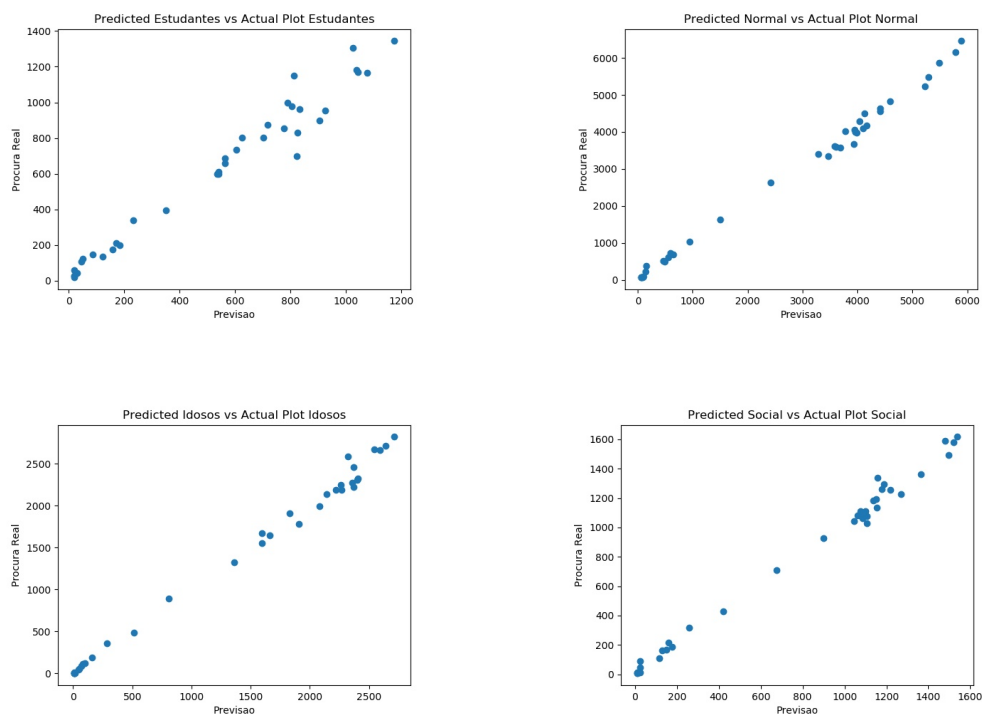


Figura C.1: Procura total para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia da semana: Terça-feira, Nebulosidade: céu limpo

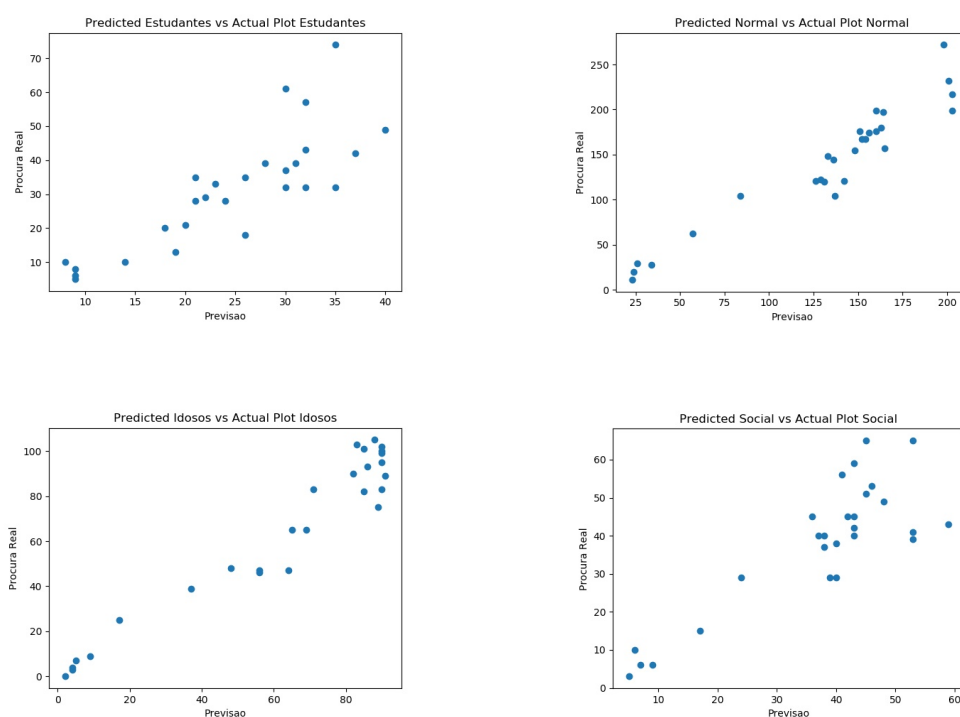


Figura C.2: Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia da semana: Terça-feira, Nebulosidade: céu limpo



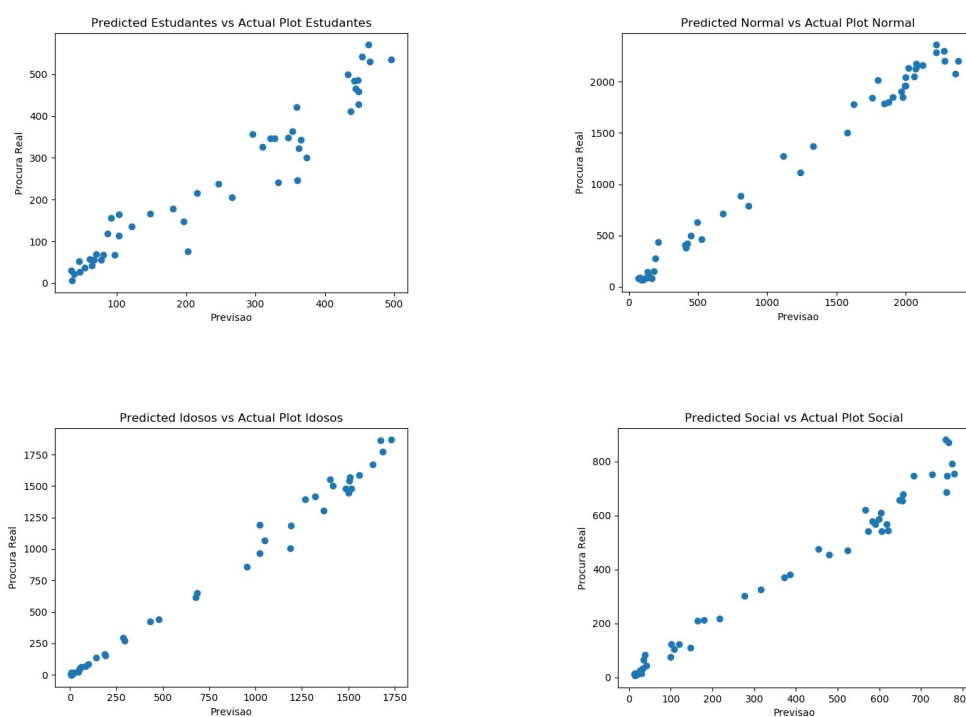


Figura C.3: Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Fim de semana: Sábado, Nebulosidade: chuva

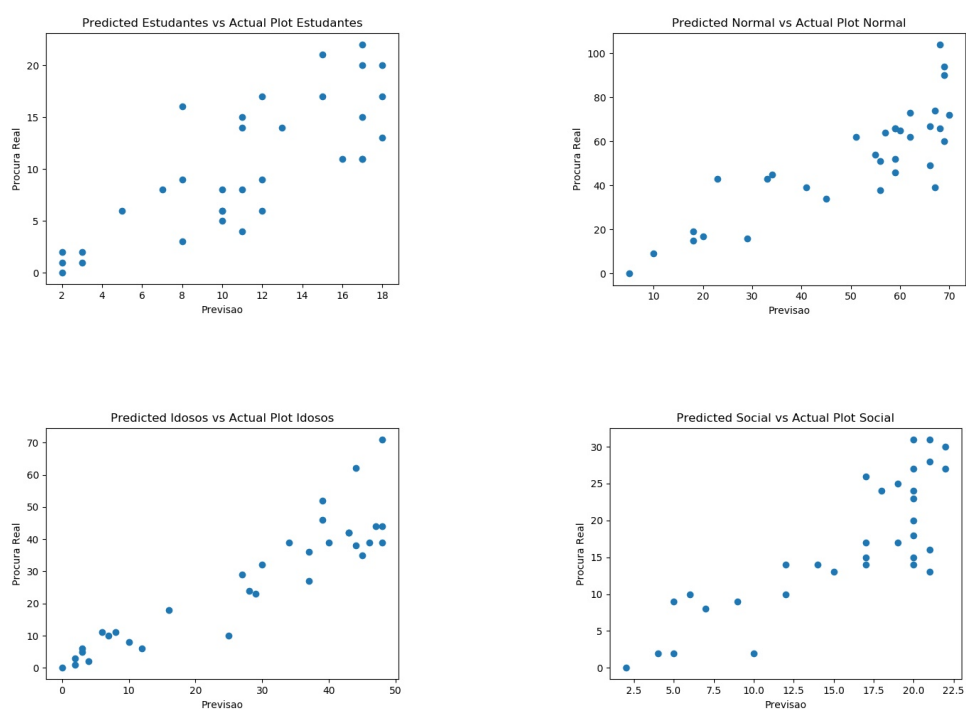


Figura C.4: Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Fim de semana: Sábado, Nebulosidade: chuva

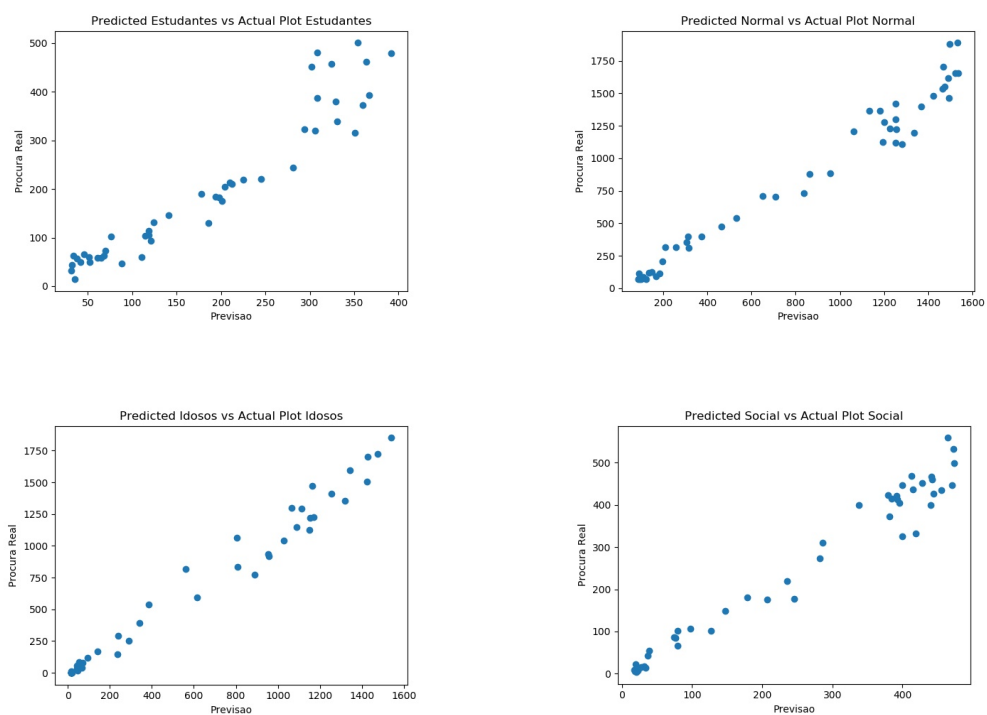


Figura C.5: Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Feriado: Quarta-feira, Nebulosidade: algumas nuvens

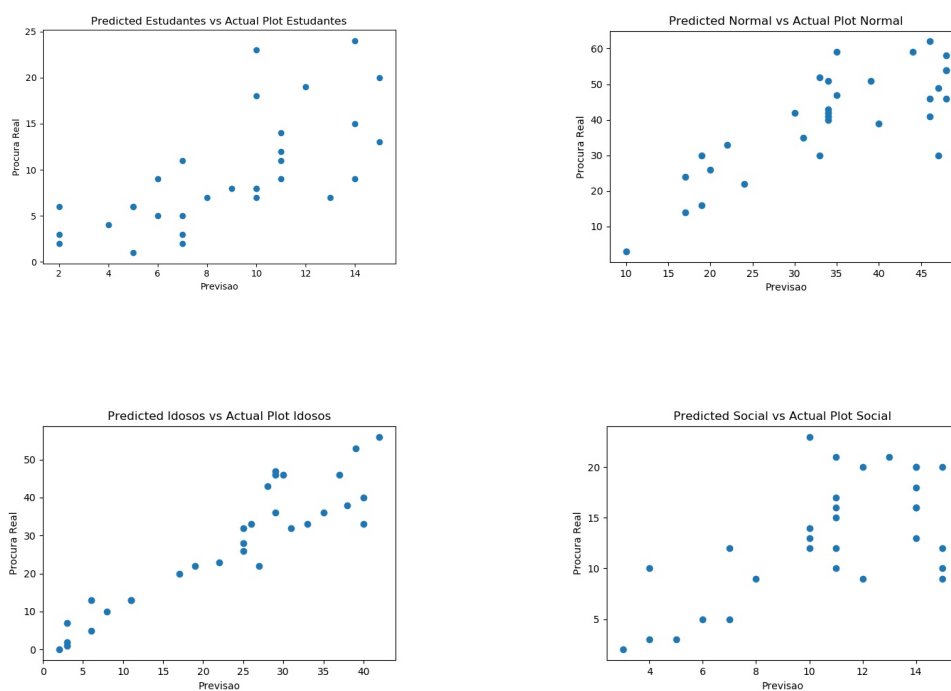


Figura C.6: Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Feriado: Quarta-feira, Nebulosidade: algumas nuvens

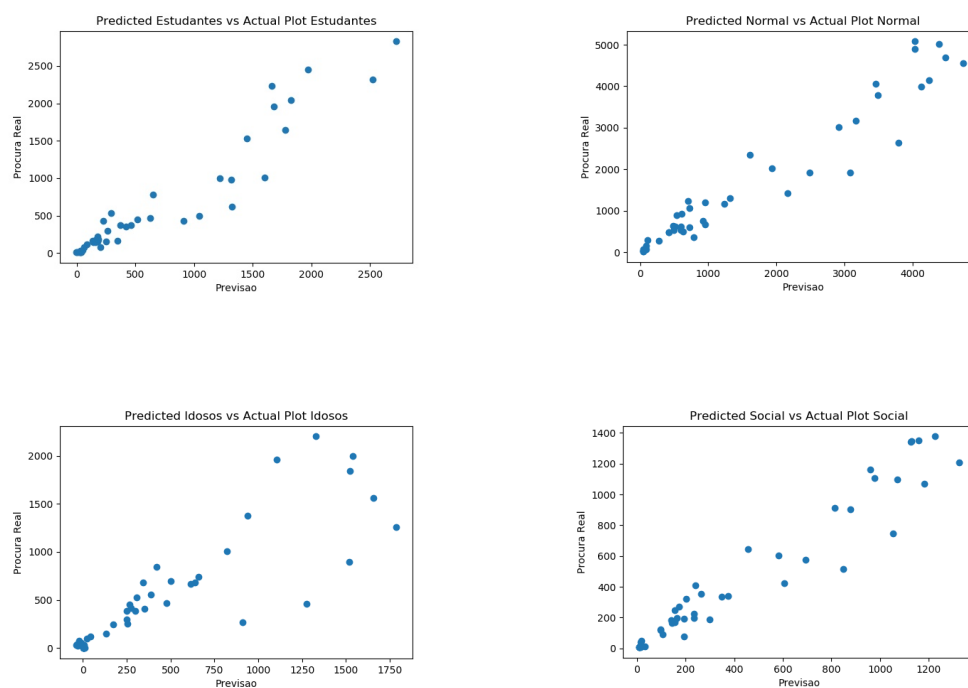


Figura C.7: Procura total de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Greve: Terça-feira, Nebulosidade: nublado

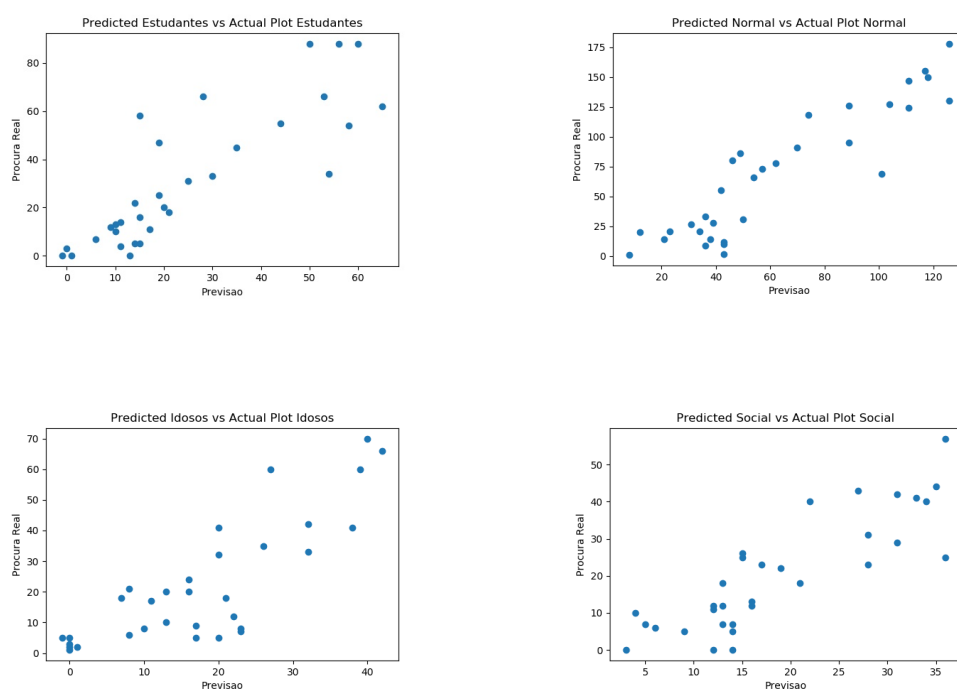


Figura C.8: Procura na linha 200 de autocarro para cada grupo de utilizador vs Procura real para cada grupo de utilizador - Dia de Greve: Terça-feira, Nebulosidade: nublado

## Anexo D

### Código da ferramenta

```
import requests
import json
import pandas as pd
from pandas.io.json import json_normalize
import numpy as np
os.environ["KERAS_BACKEND"] = "plaidml.keras.backend"
import keras
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation
from sklearn.preprocessing import MinMaxScaler
from sklearn import preprocessing
from keras import regularizers
from keras import backend
from keras import backend as K
import sklearn.metrics, math
from datetime import datetime, date, time, timedelta
from keras.models import load_model
from pickle import load
import xlwt
import xlswriter

#defini ao de funcoes necessarias ao funcionamento
def mean_abs_error(y_true, y_pred, x):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    i=0
    val=0
    while i < len(y_true):
        if y_true[i,x]!=0 and y_pred[i,x]!=0:
```

```

        if y_true[i,x]>y_pred[i,x]:
            val=val+abs(y_true[i,x] - round(y_pred[i,x]))

        if y_true[i,x]<y_pred[i,x]:
            val=val+abs(round(y_pred[i,x]) - y_true[i,x])

    i=i+1

    return round((val/len(y_true)))

#metrics
    # root mean squared error (rmse) for regression (only for Keras tensors)
def rmse(y_true , y_pred):
    return backend.sqrt(backend.mean(backend.square(y_pred - y_true), axis=-1))

# mean squared error (mse) for regression (only for Keras tensors)
def mse(y_true , y_pred):
    return backend.mean(backend.square(y_pred - y_true), axis=-1)

# coefficient of determination (R^2) for regression (only for Keras tensors)
def r_square(y_true , y_pred):
    SS_res = K.sum(K.square(y_true - y_pred))
    SS_tot = K.sum(K.square(y_true - K.mean(y_true)))
    return ( 1 - SS_res/(SS_tot + K.epsilon()) )

#####

#hora do dia em que se esta a pedir a previsao
now=datetime.now()

#pedido a API das previsoes meteorologicas para as horas seguintes ate as 24h
r="https://api.weatherbit.io/v2.0/forecast/hourly?city=Porto ,
PT&key=55133b193839429f8d862dcd4a28bf18&hours="+str(23-now.hour)
response = requests.get(r)

#transformacao do ficheiro JSON em PandasDataframe

df = json_normalize(response.json(), 'data')

#eliminacao de informacao nao necessaria

```



```
del df[ 'weather.icon' ]  
del df[ 'weather.description' ]  
del df[ 'ozone' ]  
del df[ 'solar_rad' ]  
del df[ 'wind_gust_spd' ]  
del df[ 'vis' ]  
del df[ 'snow_depth' ]  
del df[ 'clouds' ]  
del df[ 'ts' ]  
del df[ 'app_temp' ]  
del df[ 'dhi' ]  
del df[ 'dni' ]  
del df[ 'ghi' ]  
del df[ 'slp' ]  
del df[ 'snow' ]  
del df[ 'wind_dir' ]  
del df[ 'wind_cdir' ]  
del df[ 'wind_cdir_full' ]  
del df[ 'pop' ]  
del df[ 'pod' ]  
del df[ 'uv' ]  
del df[ 'clouds_low' ]  
del df[ 'clouds_mid' ]  
del df[ 'clouds_hi' ]  
del df[ 'timestamp_utc' ]  
del df[ 'datetime' ]  
del df[ 'precip' ]
```

*#tratamento de dados*

```
df[ 'timestamp_local' ]=df[ 'timestamp_local' ]. str. split( 'T', expand=True )[1]  
df[ 'timestamp_local' ]=df[ 'timestamp_local' ]. str. split( ':', expand=True )[0]  
df[ 'timestamp_local' ]=df[ 'timestamp_local' ]. astype( float )*2  
df[ 'pres' ]=df[ 'pres' ]. round()  
df[ 'wind_spd' ]=df[ 'wind_spd' ]. round()  
df[ 'wind_spd' ]=df[ 'wind_spd' ]*3.6  
df=df[[ 'timestamp_local', 'temp', 'wind_spd', 'rh', 'dewpt', 'pres',  
       'weather.code' ]]  
  
df[ 'temp' ]=df[ 'temp' ]. astype( int )
```

```

df['weather.code']=df['weather.code'].astype(int)

newdf = pd.DataFrame(np.repeat(df.values,2,axis=0))
newdf.columns = df.columns

#transformacao da hora em codigos hora em que a rede foi treinada
for i in range(1,len(newdf)):
    if (i % 2) != 0:
        newdf.at[i,'timestamp_local']= newdf['timestamp_local'].values[i]+1

newdf.set_index('timestamp_local')

#dicionario de nebulosidade
dict = {200 : 5, 201 : 5, 202 : 5,230 : 5,231 : 5,232 : 5,233 : 5,
300 : 5,302 : 5,500 : 4,501 : 4,502 : 4,511 : 4,520 : 4,521 : 4,
522 : 4,800 : 0,801 : 1,802 : 2,803 : 3,900 : 4}

#transformacao dos codigos de nebulosidade fornecidos pela API
em codigos de nebulosidade que a rede foi treinada
newdf['weather.code']= newdf['weather.code'].map(dict)

#pedidos de informa o ao utilizador
ds = input("Qual o dia da semana (Segunda:1, Ter a :2, Quarta:3, Quinta:4,
Sexta:5, Sabado:6, Domingo:7)? :")
Greve = input(" dia de greve (Sim:1, Nao:0)? :")
Feriado = input(" dia de feriado (Sim:1, Nao:0)? :")
mes=now.month
newdf['DS']=ds
newdf['Greve']=Greve
newdf['Feriado']=Feriado
newdf['mes']=mes

#transforma o das variaveis em floats
newdf=newdf.astype(float)
newdfl=newdf

#previsao total da rede

```

```

#carregamento do normalizador
min_max_scale = load(open('scaler_rede.pkl', 'rb'))
inputs = min_max_scale.transform(newdf)

#carregamento da rede neuronal de previsao total
model = load_model('modelo_rede_final.h5',
custom_objects={"rmse": rmse,"r_square": r_square,"mse": mse })

#execucao da previs o
pred=model.predict(inputs ,verbose=0,batch_size=1)

#arredondamento dos valores de procura e alteracao
dos valores negativos para zero
pred=pred.round()
pred=pred.clip(min=0)

# cria o do ficheiro excel
workbook = xlsxwriter.Workbook('PrevisaoTotal.xlsx')
worksheet = workbook.add_worksheet('PrevisaoTotal')

#escrita dos valores de previsao e respectivas horas no ficheiro excel
worksheet.write(0, 0,'Hora')
worksheet.write(0, 1,'Estudantes')
worksheet.write(0, 2,'Normal')
worksheet.write(0, 3,'Idosos')
worksheet.write(0, 4,'Social')

row = 1
for hora in (newdf['timestamp_local']):
    worksheet.write(row, 0, '{0:02.0f}:{1:02.0f}'.format(*divmod((hora/2)
    * 60, 60)))
    row += 1

row = 1
col = 1

for item in (pred):

```

```

worksheet.write(row, col, item[0])
worksheet.write(row, col + 1, item[1])
worksheet.write(row, col + 2, item[2])
worksheet.write(row, col + 3, item[3])
row += 1

#encerramento do ficheiro excel
workbook.close()

#previsao por linha da rede

#dicionario linhas
linhas=([1,17,18,22,26,48,61,69,94,98,99,103,104,106,200,201,202,203,204,
205,206,207,208,209,300,301,302,303,304,305,400,401,402,500,501,502,503,
504,505,506,507,508,600,601,602,603,604,700,701,702,703,704,705,706,707,
800,801,803,804,805,806,900,901,902,903,904,905,906,907,1.1,10.1,11.1,
12.1,13.1,3.1,4.1,5.1,7.1,8.1,9.1,94.1,95.1])

#carregamento do normalizador
min_max_scaler = load(open('scaler_linha.pkl', 'rb'))

#carregamento da rede neuronal de previsao para linhas
modell = load_model('modelo_linha_final.h5',
custom_objects={"rmse": rmse,"r_square": r_square,"mse": mse })

#criacao do excel
workbook = xlswriter.Workbook('PrevisaoPorLinha.xlsx')

#previsao por linha com os dados meteorologicos e temporais
e respectiva gravacao no ficheiro excel
#em cada previsao de linha criada uma nova folha no ficheiro excel
for linha in (linhas):
    #definicao da ordem de entrada de variaveis na rede
    newdfl['Service_ID']=linha*10
    newdfl=newdfl[['Service_ID','timestamp_local','temp','wind_spd','rh',
'dewpt','pres','weather.code','DS','Greve','Feriado','mes']]
    inputs_l = min_max_scaler.transform(newdfl)

#execu o da previsao

```

```
predl=modell.predict(inputs_1,verbose=0,batch_size=1)

#arredondamento dos valores de procura e alteracao
dos valores negativos para zero
predl=predl.round()
predl=predl.clip(min=0)

#escrita dos valores de previsao e respectivas horas
no ficheiro excel
worksheet = workbook.add_worksheet('Previsao_'+ str(linha))

worksheet.write(0, 0, 'Hora')
worksheet.write(0, 1, 'Estudantes')
worksheet.write(0, 2, 'Normal')
worksheet.write(0, 3, 'Idosos')
worksheet.write(0, 4, 'Social')

row = 1
for hora in (newdf['timestamp_local']):
    worksheet.write(row, 0, '{0:02.0f}:{1:02.0f}'.
        format(*divmod((hora/2) * 60, 60)))
    row += 1

row = 1
col = 1

for item in (predl):
    worksheet.write(row, col, item[0])
    worksheet.write(row, col + 1, item[1])
    worksheet.write(row, col + 2, item[2])
    worksheet.write(row, col + 3, item[3])
    row += 1

#encerramento do ficheiro excel
workbook.close()
```

```
#informacao ao utilizador que a ferramenta ja  
executou as previsoes  
print(' _Previsao _para _as _proximas _'+str(horas)+ ' _horas _executada .  
\n_Por _favor _consulte _os _ficheiros _gerados _com _os _nomes  
PrevisaTotal.xlsx_e _PrevisaoPorLinha.xlsx ')  
  
}
```

# Bibliografia

- AMP. Amp em números, 2019. URL [http://portal.amp.pt/pt/4/stats/#FOCO\\_4](http://portal.amp.pt/pt/4/stats/#FOCO_4).
- P. Arana, S. Cabezudo, and M. Peñalba. Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A: Policy and Practice*, 59:1–12, 2014. ISSN 09658564. doi: 10.1016/j.tra.2013.10.019.
- T. A. Badgwell, J. H. Lee, and K.-h. Liu. *Reinforcement Learning – Overview of Recent Progress and Implications for Process Control*, volume 44. Elsevier Masson SAS, 2018. ISBN 9780444642417. doi: 10.1016/B978-0-444-64241-7.50008-2.
- D. R. Baughman and Y. A. Liu. *Neural networks in bioprocessing and chemical engineering*. Academic press, 2014.
- J. Bergstra and Y. Bengio. Random Search for Hyper-Parameter Optimization. 13:281–305, 2012.
- F. Chollet. keras. <https://github.com/fchollet/keras>, 2015.
- A. Cochocki and R. Unbehauen. *Neural Networks for Optimization and Signal Processing*. John Wiley Sons, Inc., USA, 1st edition, 1993. ISBN 0471930105.
- M. Cools, E. Moons, L. Creemers, and G. Wets. Changes in Travel Behavior in Response to Weather Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2157:22–28, 2010. ISSN 0361-1981. doi: 10.3141/2157-03.
- V. Costa, T. Fontes, J. Borges, and T. Dias. Impacts of Weather Conditions in Urban Public Transport: Understanding the Effects of Climatic Changes using Big Data. (August 2016), 2017.
- L. Creemers, G. Wets, and M. Cools. Meteorological variation in daily travel behaviour: evidence from revealed preference data from the Netherlands. *Theoretical and Applied Climatology*, 120 (1-2):183–194, 2015. ISSN 14344483. doi: 10.1007/s00704-014-1169-0.
- A. Cyril, R. H. Mulangi, and V. George. Modelling and Forecasting Bus Passenger Demand using Time Series Method. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 460–466, 2018. doi: 10.1109/ICRITO.2018.8748443.

- L. de Montigny, R. Ling, and J. Zacharias. The Effects of Weather on Walking Rates in Nine Cities. *Environment and Behavior*, 44(6):821–840, 2011. ISSN 00139165. doi: 10.1177/0013916511409033.
- J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, July 2011. ISSN 1532-4435.
- Eurostat. Eurostat - estatísticas sobre o transporte de passageiros ,2019.
- B. Fields, U. o. N. Orleans, G. C. R. C. f. E. Resiliency, Transportation, R. Administration, and I. Technology. Active Transportation Measurement and Benchmarking Development: New Orleans State of Active Transportation Report 2010. page 103p, 2012.
- Flickr. Flickr - metro do porto - 2019, Jun 2019.
- Z. Guo, N. H. Wilson, and A. Rahbee. The Impact of Weather on Transit Ridership Chicago. *TRB Annual Meeting*, pages 3–10, 2007.
- Z. Guo, N. Wilson, and A. Rahbee. Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2034:3–10, 2008. ISSN 0361-1981. doi: 10.3141/2034-01.
- S. S. Haykin et al. *Neural networks and learning machines/Simon Haykin*. New York: Prentice Hall,, 2009.
- R. Hjorthol. Winter weather – an obstacle to older people’s activities? *Journal of Transport Geography*, 28:186–191, 2013. ISSN 09666923. doi: 10.1016/j.jtrangeo.2012.09.003.
- E. Hooper, L. Chapman, and A. Quinn. The impact of precipitation on speed–flow relationships along a uk motorway corridor. *Theoretical and Applied Climatology*, 117(1):303–316, Jul 2014. ISSN 1434-4483. doi: 10.1007/s00704-013-0999-5.
- IBM SPSS Statistics for Windows. Ibm corp. released 2019, 2019. URL <https://www.ibm.com/analytics/spss-statistics-software>.
- INE. Resultados Definitivos. *Censos 2011*, 26:1–41, 2012.
- IPMA. Normais climatológicas, 2019. URL <https://www.ipma.pt/pt/oclima/normais.clima/>.
- JPN. Mudança de autocarros para o dragão - jpn, 2019, Feb 2019.
- A. J. Kalkstein, M. Kuby, D. Gerrity, and J. J. Clancy. An analysis of air mass effects on rail ridership in three US cities. *Journal of Transport Geography*, 17(3):198–207, 2009. ISSN 09666923. doi: 10.1016/j.jtrangeo.2008.07.003.
- D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014.



- A. Lewin. Temporal and weather impacts on bicycle volumes. Technical report, 2011.
- J. Li, X. Li, D. Chen, and L. Godding. Assessment of metro ridership fluctuation caused by weather conditions in Asian context: Using archived weather and ridership data in Nanjing. *Journal of Transport Geography*, 66(35):356–368, 2018. ISSN 09666923. doi: 10.1016/j.jtrangeo.2017.10.023.
- Z. Ma, J. Xing, M. Mesbah, and L. Ferreira. Predicting short-term bus passenger demand using a pattern hybrid approach. *TRANSPORTATION RESEARCH PART C*, 39:148–163, 2014. ISSN 0968-090X. doi: 10.1016/j.trc.2013.12.008.
- A. L. Maas and A. Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. 28, 2013.
- M. K. E. Mahrsi. A Novel Method of Multi-Information Acquisition for Electromagnetic Flow Meters. page 9, dec 2014. ISSN 1424-8220. doi: 10.3390/s16010025.
- N. Mukai and N. Yoden. Taxi Demand Forecasting Based on Taxi Probe. 2012.
- Nielsen and M. A. Neural networks and deep learning, Jun 2019.
- Orlanski. A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, 56(5):527–530, 1975. ISSN 00030007-15200477.
- J. Parkin, M. Wardman, and M. Page. Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation*, 35:93–109, 02 2008. doi: 10.1007/s11116-007-9137-5.
- Porto. Terminal do bom sucesso. - portal de notícias do porto. - 2018, 2018. URL <http://www.porto.pt>.
- M. Pumperla. Hyperas. <https://github.com/maxpumperla/hyperas>, 2019.
- C. R. Rao. *Linear models and generalizations: least squares and alternatives*. Springer, 2008.
- J.-P. Rodrigue, C. Comtois, and B. Slack. *The geography of transport systems*. Routledge, 2017.
- Z. Rowland and J. Vrbka. Using artificial neural networks for prediction of key indicators of a company in global world. In T. Klieštík, editor, *Globalization and its socio-economic consequences, 16th international scientific conference proceedings, PTS I-V*, pages 1896–1903, Žilina, Slovensko, 2016. Žilina, Slovensko. ISBN 978-80-8154-191-9.
- M. Sabir, J. van Ommeren, M. J. Koetse, and P. Rietveld. Impact of weather on daily travel demand. 2010.
- P. Schmiedeskamp and W. Zhao. Estimating Daily Bicycle Counts in Seattle, Washington, from Seasonal and Weather Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2593(1):94–102, 2016. ISSN 0361-1981. doi: 10.3141/2593-12.

- K. Shaaban and D. Muley. Investigation of weather impacts on pedestrian volumes. 2016. ISSN 23521465. doi: 10.1016/j.trpro.2016.05.047.
- A. Singhal, C. Kamga, and A. Yazici. Impact of weather on urban transit ridership. *Transportation Research Part A: Policy and Practice*, 69:379–391, 2014. ISSN 09658564. doi: 10.1016/j.tra.2014.09.008.
- V. Stover and E. McCormack. The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, 15(1):95–110, 2015. ISSN 1077-291X. doi: 10.5038/2375-0901.15.1.6.
- S. Tao, J. Corcoran, M. Hickman, and R. Stimson. The influence of weather on local geographical patterns of bus usage. *Journal of Transport Geography*, 54:66–80, 2016. ISSN 09666923. doi: 10.1016/j.jtrangeo.2016.05.009.
- S. Tao, J. Corcoran, F. Rowe, and M. Hickman. To travel or not to travel: ‘Weather’ is the question. Modelling the effect of local weather conditions on bus ridership. *Transportation Research Part C: Emerging Technologies*, 86(April 2017):147–167, 2018. ISSN 0968090X. doi: 10.1016/j.trc.2017.11.005.
- TIP - Transportes Intermodais do Porto. Relatório de Contas 2013. Technical report, Porto, Portugal, 2013.
- Vertex.AI. Plaidml, 2019. URL <https://github.com/plaidml/plaidml>.
- Weatherbit.io, 2020. URL <https://www.weatherbit.io>.
- R. Xue, D. J. Sun, and S. Chen. Short-Term Bus Passenger Demand Prediction Based on Time Series Model and Interactive Multiple Model Approach. 2015(i), 2015.
- S. Yagi and A. Mohammadian. Policy simulation for New BRT and area pricing alternatives using an opinion survey in Jakarta. *Transportation Planning and Technology*, 31(5):589–612, 2008. ISSN 03081060. doi: 10.1080/03081060802087676.
- M. Zhou, D. Wang, Q. Li, Y. Yue, W. Tu, and R. Cao. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies*, 75:17–29, 2017. ISSN 0968090X. doi: 10.1016/j.trc.2016.12.001.
- J. Zupan and J. Gasteiger. *Neural Networks for Chemists: An Introduction*. John Wiley Sons, Inc., USA, 1993. ISBN 3527286039.